

THE ONLINE REGULATION SERIES 3.0

THE HANDBOOK

tech
against
terrorism





ABOUT TECH AGAINST TERRORISM

Tech Against Terrorism supports online service providers in countering terrorist use of the internet whilst safeguarding human rights. We do so by equipping service providers with the knowledge and tools required to develop resilient counterterrorism responses.

Our research shows that terrorist groups - both jihadist and far-right terrorists - consistently exploit smaller tech platforms when disseminating propaganda. At Tech Against Terrorism, our mission is to support smaller tech companies in tackling this threat whilst respecting human rights and to provide companies with practical tools to facilitate this process. As a public-private partnership, the initiative works with the United Nations Counter Terrorism Executive Directorate (UN CTED) and has been supported by the Global Internet Forum to Counter Terrorism (GIFCT) and the governments of Spain, Switzerland, the Republic of Korea, and Canada.

techagainstterrorism.org
contact@techagainstterrorism.org



TECH AGAINST TERRORISM'S POLICY ADVISORY AND RESPONSE

Tech Against Terrorism equips and enables online service providers to counter terrorist use of the internet. This 2nd edition of the Online Regulation Series has been compiled by Tech Against Terrorism's Policy Advisory and Response (PAR) team.

PAR's mission is to promote policy and legal responses to terrorist use of the internet that safeguard human rights and uphold online security.

We support online services providers (OSPs) in navigating the complexities of the legal framework around terrorist use of the internet and harmful online content. We help tech companies understand how legal requirements may impact their approach to counterterrorism and content moderation approach, and provide practical support to OSPs in adapting to new regulations.

We advocate for legal responses to terrorist use of the internet that are evidence-based and consider the diversity of the tech sector, including the variety of online services offered and the limited resources of smaller and newer platforms. We regularly engage with policymakers to provide evidence on terrorist use of the internet and existing counterterrorism efforts from the tech sectors, and advise on regulatory proposals and their implementation.

You can find more about Tech Against Terrorism's PAR's work at <https://bit.ly/3IPQG5J>
You can find more about regulatory work on our Knowledge Sharing Platform:
ksp.techagainstterrorism.org



CONTENTS

ABOUT TECH AGAINST TERRORISM	02
BACKGROUND TO THE ONLINE REGULATION SERIES	07
What is the Online Regulation Series?	07
Background and Evolution of the Online Regulation Series	07
METHODOLOGY	10
KEY TRENDS AND TECH AGAINST TERRORISM’S RECOMMENDATIONS	14
TECH AGAINST TERRORISM’S REGULATORY GUIDELINES	17
SECTION 1. LEGISLATIVE OVERVIEW	20
SECTION 2. COUNTRY ANALYSES	32
AUSTRALIA – update	34
AUSTRIA – update	37
EUROPEAN UNION – update	42
INDIA – update	47
KYRGYZSTAN – update	51
NEW ZEALAND – update	54
SINGAPORE – update	58
UNITED ARAB EMIRATES – update	61
UNITED KINGDOM – update	65
UNITED STATES – update	68
GENERAL update	72
SECTION 3. CROSS-SECTOR POLICY INITIATIVES	82
EVALUATING THE IMPACT OF THE CHRISTCHURCH CALL TO ACTION	84
BIBLIOGRAPHY	88



BACKGROUND TO THE ONLINE REGULATION SERIES

What is the Online Regulation Series?

With the Online Regulation Series, Tech Against Terrorism analyses over 100 pieces of legislation from 30 jurisdictions around the world. In doing so, we make sense of the complex regulatory landscape by which platforms are required to prevent the dissemination of illegal and harmful content on their services.

We monitor the evolution of regulations impacting online content, and we provide an overview of the legal requirements that can impact online service provider efforts to counter terrorism and violent extremism. Transparency and human rights are at the core of Tech Against Terrorism's work, and the Online Regulation Series also considers requirements related to transparency and accountability in content moderation.

The Online Regulation Series also includes Tech Against Terrorism's commentary on the regulations reviewed, assessing whether they meet their stated aims of countering illegal online content and terrorist use of the internet, as well as if they present risks to human rights and tech sector diversity. We also commend legislation that gives consideration to the rule of law, due process, and providing the necessary human rights safeguards.

The Online Regulation Series is equally addressed at tech companies, policymakers, counterterrorism experts, and all those interested in the following questions:

- How is the global regulatory landscape evolving?
- What are the implications for platforms' online counterterrorism efforts?
- How can we inform efficient and rights-safeguarding legal responses to terrorist use of the internet?

Background and Evolution of the Online Regulation Series

Policymakers' increased interest in regulating online content has led to a complex and multifaceted legal landscape, requiring platforms to prevent the dissemination of illegal and harmful content on their services. In 2017, Germany passed the Network Enforcement Act (NetzDg) and became one of the first countries to require platforms to prevent the spread of illegal and harmful content online, introducing a 1-hour removal deadline for terrorist content.¹ This marked a turning point in online regulation, which was followed by a global wave of regulatory discussions around content governance and the removal of illegal or harmful content.

Whilst the stated aims of countering harmful content are legitimate, this multiplication of regulation creates a fragmented legal landscape contradicting the global nature of the online sphere by requiring platforms to comply with a multitude of concurrent and presently conflictual legal obligations across jurisdictions.

¹ Tech Against Terrorism (2021), [The Online Regulation Series Handbook](#), p.91

This fast-changing regulatory landscape prompted Tech Against Terrorism to support tech companies in understanding these new legal requirements. We launched the Online Regulation Series in 2020, and have since analysed over a hundred laws and legislative proposals that cover or affect terrorist use of the internet. The Online Regulation Series is our response to the ongoing policy discussions around online content governance, which places the onus of countering terrorist and harmful online content on tech companies without providing them with adequate support to tackle this threat.

You can find all the editions of the Online Regulation Series, including individual blogposts and compiled Handbook, as well as a regulatory table with links to all regulations covered since 2020 on our dedicated Spotlight on our Knowledge Sharing Platform:
ksp.techagainstterrorism.org/online-regulation-spotlight/

With the first edition of the Online Regulation Series, in 2020, we covered 17 jurisdictions who had the most influential impact on the global online regulation landscape. The 2020 Online Regulation Series took the form of a series of analysis published on our website and was subsequently adapted into the first edition of our Online Regulation Series Handbook.

For the Online Regulation Series 2.0, in 2021, we expanded the scope of countries covered, focusing on Sub-Saharan Africa and Europe.

In 2022, we continued to track the evolution of the regulatory landscape and focused on providing updates to countries previously covered, whilst expanding our analysis to cover the evolution of tech platforms' counterterrorism responses since 2020. The result of our 2020 analysis is compiled in this Online Regulation Series 3.0 Handbook. Individual analysis for all the countries covered will also be available on Knowledge Sharing Platform.



tech against terrorism



METHODOLOGY



METHODOLOGY

This Online Regulation Series 3.0 focuses on changes to the online regulatory landscape between November 2021 and December 2022. We seek to address the following questions:

- Which legislative proposals, previously analysed in the Online Regulation Series, have been signed into laws since 2021?
- Which countries, not previously covered in the Online Regulation Series, have introduced or passed laws related to online content and/or terrorist use of the internet since 2021?
- What are the key trends in online regulation?

The majority of the research for the creation of this Third Series was conducted between September 2022 and November 2022. All jurisdictions covered in previous editions of the Online Regulation Series were assessed to determine whether there had been any significant regulatory changes within the allotted time-period, including new legislative proposals, revised draft laws and bills passed with different requirements.

Selection Criteria

The Online Regulation Series 3.0 features dedicated sections for those countries previously analysed by us and which have since had significant updates to their regulatory framework. In our general updates section, we cover those countries which passed laws without significant change to the draft previously covered. The general updates section also covers countries that have passed new laws which aim to regulate the online space but do not specifically concern countering terrorism.

Additional countries not previously featured in the Online Regulation Series are now incorporated based on whether there had been any new or significant pieces of online regulation. They merit inclusion because they show similarities with other regulatory frameworks analyses in the Series. Core to our selection criteria, we considered the direct implications for online service providers in countering terrorist use of the internet and moderating online content.

As online regulation continues to change globally, Tech Against Terrorism will strive to provide regular updates on the implications for tech companies, and their efforts in countering terrorist use of the internet whilst respecting human rights.



JURISDICTIONS COVERED BY THE ORS 3.0





KEY TRENDS AND TECH AGAINST TERRORISM'S RECOMMENDATIONS

By analysing key regulatory trends, we are able to outline best practices for policymakers to create transparent legal responses to terrorist use of the internet. Crucially, we focus on identifying those provisions presenting risks for human rights and tech sector diversity. In 2022, we identified five global key trends in regulating harmful and illegal content online:

- Short-term removal deadlines requiring platforms to act on reported content within a limited timeframe
- Lack of definitional clarity around what constitutes terrorist content
- Broad transparency requirements unsuited to platforms' specificities and capacity
- Backlash against platforms' capacity to remove content based on their own guidelines
- Lack of public consultation and lack of transparent regulatory processes

Based on the trends identified, we outline the following recommendations for policymakers to consider as regulation continues to evolve:

- **Reconsider strict removal deadlines.** We note an increase in legal provisions media and content hosting platforms face this burden in particular. Removal deadlines typically range between 1 to 24 hours upon notification of the content to the Online Service Providers (OSPs), with some extending to 72 hours.² Certain regulations, for instance in Australia, consider terrorist content as high priority content for removal, with a shorter removal deadline.
 - ▶ Tech Against Terrorism cautions against strict removal deadlines that do not consider the capacity of smaller and newer OSPs to comply. It is also inadvisable to penalise service providers for lacking the resources to comply.
 - ▶ Tech Against Terrorism also warns of the unintended consequences of imposing short removal deadlines, especially when they do not allow sufficient time to assess the legality of content. Platforms risk erring on the side of over-removal, including the removal of legal content, to avoid penalties.
- **Improve definitional clarity of terrorist and illegal content.** The majority of bills and laws analysed in this Series continue to be characterised by a lack of definitional clarity around what constitutes illegal speech, and around terrorist content in particular. This risks the removal of legitimate content caught by definitional uncertainties. Mirroring this, there is also the risk of regulation failing to account for the realities of content used by terrorists and violent extremist actors. Tech Against Terrorism finds that terrorists and violent extremist actors exploit the limit of what constitutes legal content to evade content moderation via so-called "borderline" or "grey area" content.³ Correctly identifying terrorist and violent extremist content thus represents a significant challenge for tech companies.

²As with the IT Rules in India.

³ Whilst "borderline" or "grey area" content remains an umbrella term requiring further definitional clarity, Tech Against Terrorism consider such content via the prism of the tactical used by terrorist and violent extremist actors to evade content moderation and deplatforming. You can learn more about Tech Against Terrorism's position on such content in our dedicated podcast episode on "[Sanitising Extremism: 'Borderline Content' and Antisemitism Online](#)".

- ▶ Tech Against Terrorism encourages policymakers to provide clear frameworks and guidance for tech companies to correctly and swiftly identify terrorist content on their services.
- ▶ Tech Against Terrorism also recommends that policymakers support initiatives aimed at strengthening OSPs' understanding of the online threat landscape in order to build capacity to detect terrorist counter strategies to evade moderation. Our Terrorist Content Analytics Platforms (TCAP) provides support to 100+ OSPs to swiftly detect verified terrorist content, and our Policy Advisory and Response Team supports over fifty service providers to build both their understanding of and response to the online threat.
- Require transparency reporting from tech companies in accordance with their size, context and capabilities. Provisions requiring platforms to publish regular transparency reports continue to be included in the majority of online regulation we have analysed. These provisions typically require platforms to publish metrics on their content moderation practices and compliance with local laws, including compliance with removal orders received from competent authorities. These provisions often contain detailed specifications of which metrics are to be included in the transparency reports.
- ▶ Tech Against Terrorism commends the regulatory focus on transparency reporting. However, we call on policymakers to refrain from mandating set transparency reporting templates which ignore the specificities and capacities of each platform. Instead, we call on policymakers to encourage the use of Tech Against Terrorism's Transparency Reporting Guidelines to improve transparent and accountable reporting, from governments and tech companies alike, on the key metrics applicable to online counterterrorism activity.⁴
- Empower platforms to act on their own guidelines. With the inclusion of proposed legislation from Poland and Brazil in the Online Regulation Series 2.0, Tech Against Terrorism has noted a schism in online regulation, whereby regulations can be divided between those aimed at encouraging increased moderation of online content, and regulation aimed at restricting platforms' capacity to act on their own community guidelines. Proposals and laws to limit OSPs' acting on their own rules have emerged, for instance, in Poland, Brazil and Texas.⁵ Whilst regulations in the second category of regulation are for the most part in the proposal stage or otherwise being currently debated, provisions limiting platforms' moderation abilities risk further fragmenting the regulatory landscape as platforms would have to comply with contradictory requirements around content removal across jurisdictions (at times within the same country, as is the case in the United States).
- ▶ Platforms' community guidelines often go beyond what is legally required to tackle harmful content, and represent a critical tool for platforms to quickly adapt their moderation practices to the evolution of the online threat. Restraining platforms' ability to act on their community guidelines risks leading to terrorist and violent extremist actors' increased use of borderline content, chiefly propaganda content within the remit of legality.

⁴ Tech Against Terrorism (2021), [Transparency Reporting Guidelines on Online Counterterrorism Efforts](#).

⁵ For Poland and Brazil, please see the dedicated entries in the [Online Regulation Series 2.0](#). For Texas, please see the US entry in this handbook, p.

- Engage in public consultations. The public consultations on proposed online regulation conducted in Canada and Singapore highlight the principal obstacles to creating legal responses to terrorist use of the internet that are scalable, capable of safeguarding human rights, and accounting for the diversity of the tech sector. The consultation on the Singaporean Online Safety Bill, is also interesting in highlighting how platforms' moderation practices are understood by users.
- ▶ Tech Against Terrorism recommends that all regulatory proposals are accompanied by public consultation processes, and for the results of the consultations to be made publicly available by governments.



TECH AGAINST TERRORISM'S REGULATORY GUIDELINES

Tech Against Terrorism has established these guidelines to improve global regulation of terrorist and violent extremist use of the internet. In our analysis of the growing number of legal provisions around the world, the majority of regulations seeking to tackle content are unlikely to be effective in countering the terrorist threat, and risk adversely affecting global digital rights and tech sector innovation.

Our Guidelines are meant to inform global regulatory developments as policymakers continue to develop legislation requiring platforms to act as front-line defenders against terrorist and violent extremist content. We have created these guidelines based on the principal findings of the Online Regulation Series, and on insights gained by providing direct counterterrorism and violent extremist support to over fifty tech companies across the tech eco-system.

Our Guidelines aim to encourage resilient legal responses to terrorist use of the internet, in consideration of adversarial risks and of challenges encountered by small and medium tech companies in countering the dissemination of terrorist and violent extremist content.

All regulation aiming to counter terrorist and violent extremist use of the internet should abide by the following principles:

1. **Evidence base:** Regulation – including specific provisions – should be justified by a clear basis in evidence.
2. **Purpose alignment:** Governments should ensure that regulatory provisions will, based on available data and evidence, be conducive to meeting the stated aims of the regulation.
3. **Risk assessment:** Governments should conduct and publish risk assessments of their regulations and specific regulatory provisions. Assessments should be made across areas such as:
 - a. Probability of threat actor adversarial shift and/or displacement (as opposed to disruption) of terrorist activity online.
 - b. Risks to human rights and fundamental freedoms, including digital rights and freedom of expression online and the potential entrenchment of existing societal biases, including discrimination against minority and/or vulnerable groups.
 - c. Potential unintended negative consequences: these include (but are not limited to: i) incentivisation of increased use of non-transparent content moderation solutions, including automated removal mechanisms or industry collaborative schemes; ii) the removal of key evidence material; iii) the replication of the law by non-democratic states.
 - d. Impact on tech sector competition, innovation, and diversity.

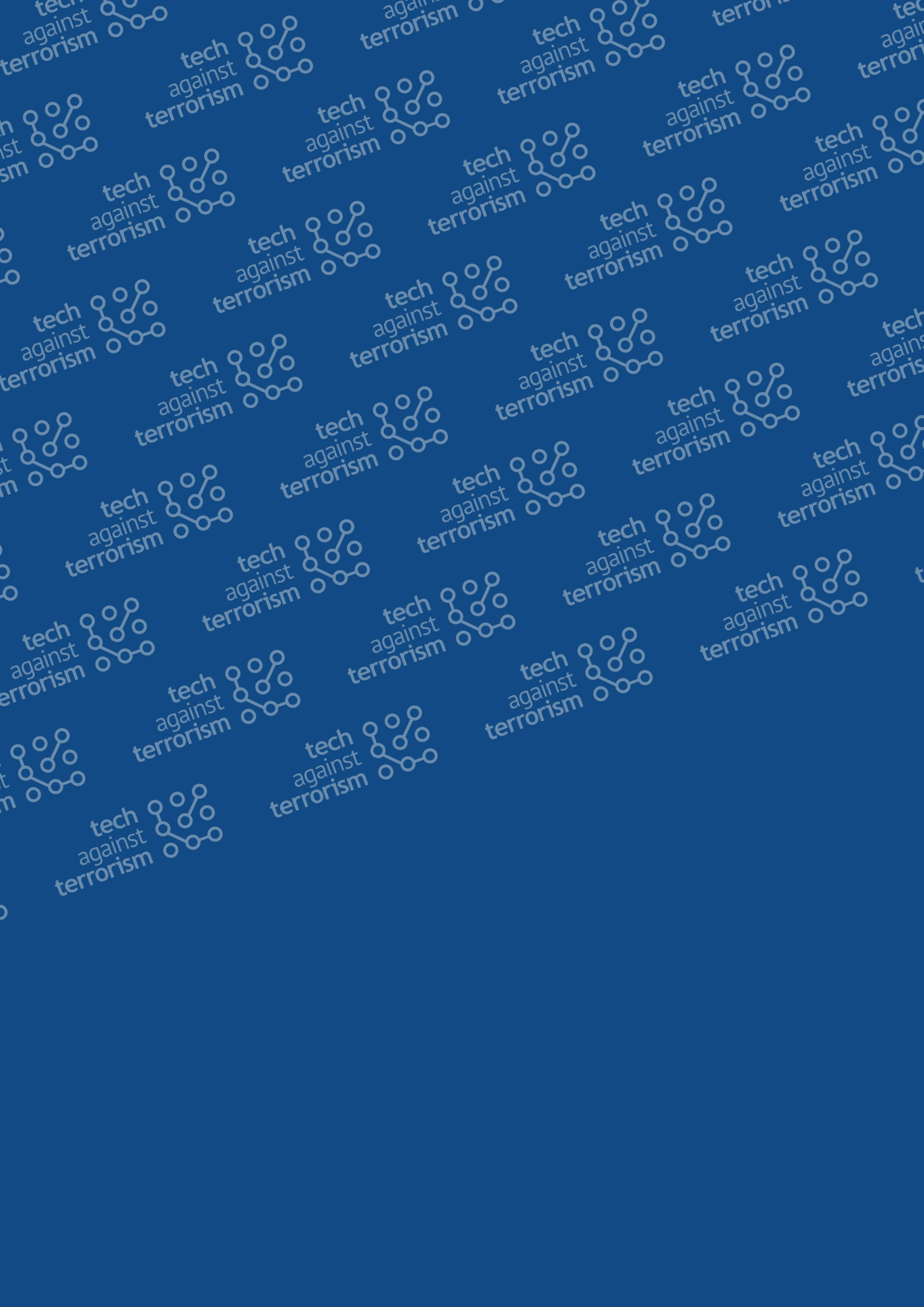
4. **Open consultation:** Governments should conduct open and transparent consultation processes and seek direct input from counterterrorism experts, digital rights advocates, and civil society organisations. A comprehensive summary of the responses received should also be published, along with any potential regulatory changes governments make as a result.
5. **Proportionality:** Governments should ensure that regulatory provisions set realistic expectations for platforms of all sizes affected by the regulation and do not disproportionately punish smaller and newer platforms.
6. **Operability:** Governments should ensure that definitions are clear and narrow, and that provisions can be operationalised by all platforms in scope of the regulation.
7. **Legal certainty:** Governments should ensure that effects of compliance and/or non-compliance are predictable and that platforms have legal certainty.
8. **Rule of law:** Governments should only mandate removal of illegal content and speech, and should avoid the promotion of extra-legal mechanisms such as the outsourcing of adjudication of illegality to platforms.
9. **Human rights and freedom of expression:** Governments should ensure that regulation complies with international norms around human rights and freedom of expression, including international human rights law. Special considerations should be given to principles developed by global civil society, such as the [Manila Principles](#) and the [Santa Clara Principles](#), and regulation should avoid introducing measures which risks undermining these and general human rights standards, including intermediary liability protections.
10. **Process vs outcome:** Regulation should adequately balance procedural and outcome targets. Consistently accurate moderation of online content at scale is impossible. Regulation should account for this by placing adequate focus on both process and outcome. Governments should also consider and encourage solutions that look beyond the removal – leaving up dichotomy.⁶
11. **Safeguards:** Regulation should have strong safeguards in place to ensure that it is not abused to censor legal and legitimate content online, and governments should monitor and mitigate against this risk, for example via redress and appeal mechanisms.
12. **Government transparency:** Governments should – in addition to producing [transparency reports on their online counterterrorism efforts](#) – be transparent about how they are implementing regulation and produce detailed transparency reports about its implementation.
13. **Evaluation:** Governments should commission independent reviewers to help assess whether the law has a) been effective in achieving its stated aims b) led to any undue restriction of legal or otherwise legitimate content and speech.

⁶ Governments and regulators should be realistic about outcome targets and avoid introducing regulation that assumes that removing all terrorist content without any impact on legal and otherwise legitimate content is possible. Should governments choose to introduce regulation that mandates the removal of terrorist content, they should – in line with the risk assessments mentioned above – be transparent about what potential errors and risks to legitimate speech may be associated with such mechanisms. Governments and regulators should not ignore procedural aspects. Whilst avoiding mistakes in regulatory compliance with regard to content removal is difficult, governments can support improved clarity and transparency around how content moderation decisions are made and enforced via regulation.

How will the Guidelines be used?

In future assessment of online regulations, Tech Against Terrorism will use the Guidelines to review new and emerging regulation, which will be assessed against the following criteria based on the Guidelines:

1. **Evidence base:** Does the regulation provide an evidence basis that justifies its introduction and specific provisions?
2. **Purpose alignment:** Do we assess that the regulation will be effective in meeting its stated objectives?
3. **Risk assessment:** Has the government introducing the law carried out a risk assessment across the four priority areas?
4. **Open consultation:** Has an open consultation process with input from counterterrorism and digital rights advocates taken place? Have the results been published, along with any potential changes made to the regulation as a result of the feedback provided?
5. **Proportionality:** Is the regulation realistic and proportionate in its expectations on smaller platforms?
6. **Operability:** Are definitions clear and narrow? Are provisions likely to be operable for all platforms in scope of the regulation?
7. **Legal certainty:** Is it our assessment that platforms will have legal certainty when complying with the regulation?
8. **Rule of law:** Is the regulation compliant with the rule of law?
9. **Human rights and freedom of expression:** Is it our assessment that the regulation will protect human rights and freedom of expression online?
10. **Process vs outcome:** Does the regulation strike an appropriate balance between process and outcome focussed targets?
11. **Safeguards:** Does the regulation have safeguards in place to prevent abuse of the law for political purposes?
12. **Transparency reporting:** Does the government publish regular transparency reports on online counterterrorism efforts and removal requests sent to tech companies?
13. **Evaluation:** Has the government commissioned an independent review of the regulation?





SECTION 1.

LEGISLATIVE OVERVIEW

With this third edition of the Online Regulation Series, we focus on providing updates to jurisdictions analysed in the first and second editions, meaning that for each jurisdiction included in this handbook we opted to focus our analysis on newly introduced or enacted legislation. To ensure a complete overview of the jurisdiction covered, this first section of the Handbook provides a high-level overview of the regulatory landscape in all jurisdictions covered in the Online Regulation Series 3.0.

For a more detailed understanding of the legal requirements for tech companies, please see the dedicated entries for each of the jurisdictions covered.

Editorial note: The legislative overview is reflective of the legal framework and state of regulatory discussions as of December 2022, unless stated otherwise.



EUROPEAN UNION

For our most recent analysis of the EU legal framework, please see the dedicated section [p.44](#). Tech Against Terrorism had previously analysed the EU legal framework in the Online Regulation Series [1.0](#) and [2.0](#)

EU Legislative framework overview – Insights from previous ORS editions

Legal framework

[Digital Services Act, 2022](#)

- Transparency reporting obligations
- Requirements for Very Large Online Platforms (VLOPs), including crisis response mechanisms

[Regulation on preventing the dissemination of terrorist content online \(TCO\)](#)

- 1-hour removal deadline for terrorist content, upon receipt of a removal order from a competent authority
- Introduction of ““specific measures” to prevent terrorist content if instructed by competent authorities”
- Preservation of removed terrorist content for six months
- Transparency reporting obligations

[EU Security Union Strategy, 2020](#)

[EU Counter Terrorism Agenda, 2020](#)

[Directive \(EU\) 2017/541 on combatting terrorism, 2017](#)

[EU Code of Conduct on Illegal Hate Speech, 2016](#)

[EU Crisis protocol, 2015](#)

[European Agenda on Security, 2015](#)

[EU Counter Terrorism Strategy, 2005](#)

Companies have the possibility to participate in several voluntary collaborative schemes together with European law enforcement agencies and Member States.



SINGAPORE

For our most recent analysis of Singapore’s legal framework, please see the dedicated section [p.58](#). Tech Against Terrorism has previously analysed Singapore’s legal framework in the Online [Regulation Series 1.0](#).

Singapore Legislative framework overview – Insights from previous ORS editions

Legal framework

Protection of Online Falsehoods and Manipulation Bill (POFMA), 2019

- Correction notices and requests to disable access from government ministers, failure to comply can result in financial penalties of up to S\$1,000,000 per day

Internet Code of Practice, 2016

- All internet content and service providers operating in Singapore must comply
- Prohibition of “objectionable” material, on the grounds of public interest, public morality, public order, public security, national harmony or is otherwise prohibited by applicable Singapore laws



NEW ZEALAND

For our most recent analysis of New Zealand’s legal framework, please see the dedicated section [p.54](#). Tech Against Terrorism had previously analysed New Zealand’s legal framework in the [Online Regulation Series 2.0](#).

New Zealand Legislative framework overview – Insights from previous ORS editions

Legal framework

[Aotearoa Code of Practice, 2022](#)

- Signatories agree to four commitments:
 - Reduce the prevalence of harmful content online
 - Empower users to have more control and make informed choices
 - Enhance transparency of policies, processes and systems
 - Support independent research and evaluation
- Annual public compliance reports
- Complaint mechanism for New Zealand residents if they believe a signatory is not honouring its commitments
- Removal of signatories for breaches of the code

[Content Regulatory Review, 2021](#)

[Counter-Terrorism Legislation Act, 2021](#)

[Urgent Interim Classification of Publications Act, 2021](#)

- Takedown issued by an Inspector of Publications for objectionable content.
- Refusal or failure to take down content can result in court orders and fines

[Prevention of Online Harm Act, 2021](#)

[Terrorism Suppression \(Control Orders\) Act, 2019](#)

[Harmful Digital Communications Act \(HDCA\), 2015](#)

- Exemption from legal liability for user-generated content, as long as the platform is in compliance with Article 24 (notifying the author, or removing/disabling content, within 48 hours)

[Search and Surveillance Act, 2012](#)

[Terrorism Suppression Act, 2002](#)

[Film, Videos and Publications Classifications Act, 1993](#)



UNITED KINGDOM

For our most recent analysis of the UK legal framework, please see the dedicated section [p.65](#). Tech Against Terrorism had previously analysed the UK 's legal framework in [the Online Regulation Series 1.0](#).

United Kingdom Legislative framework overview – Insights from previous ORS editions
Legal framework
Draft Online Safety Bill , presented to Parliament in 2021
<p>Interim Code of Practice on Terrorist Content and Activity Online, 2020</p> <p>Tech companies that host user-generated content should comply with 5 principles:</p> <ul style="list-style-type: none"> • Identify and prevent terrorist use of the internet (TUI) in the UK • Minimise potential search results for terrorist content • Partake in cross-industry collaborations to find solutions to TUI • Implement user reporting, complaints and redress mechanisms • Support UK authorities in investigating and prosecuting terrorist offences
<p>Minimum harms “small and medium size enterprises” should consider</p> <ul style="list-style-type: none"> • Targeted radicalisation of vulnerable users • Sharing of terrorist content, including propaganda • Posting of URLs to terrorist content • Live broadcast of terrorist activity
<p>Interim Approach for Regulating Video-Sharing Platforms (VSPs), 2020</p> <ul style="list-style-type: none"> • VSPs must protect users under age 18 from accessing restricted material, and must implement user appeal and independent redress mechanisms. • Ofcom can serve enforcement notices and financial penalties for breaches of compliance
Counter-Terrorism and Border Security Act , 2019
Terrorism Act , 2006
Terrorism Act , 2000



 **AUSTRALIA**

For our most recent analysis of Australia’s legal framework, please see the dedicated section [p.34](#). Tech Against Terrorism had previously analysed Australia’s legal framework in the Online Regulation Series [1.0](#) and [2.0](#).

Australia Legislative framework overview – Insights from previous ORS editions

Legal framework

[Online Safety Act, 2021](#)

- 24-hour removal deadline upon notification by the e-Safety Commissioner.
- Financial penalties for non-compliance
- Mandatory reporting requirements
- Platform responsibility to ensure users are protected from harmful content on their platforms

[Surveillance Legislation Amendment \(Identify and Disrupt\) Act, 2021](#)

[Telecommunications Legislation Amendment \(International Production Orders\) Act, 2020](#)

[Criminal Code Amendment \(Sharing of Abhorrent Violent Material\) Act, 2019](#)

Violations can be sanctioned by:

- A fine of up to \$1.5m or up to three years in prison for an individual providing the content or hosting services
 - A fine of up to \$7.5m or 10% of annual revenue for each offence (for a company)
- The e-Safety Commissioner can investigate platforms, enforce actions, and block access in Australia to content hosted overseas

[Telecommunications and Other Legislation Amendment \(Assistance and Access\) Act, 2018](#)

[Enhancing Online Safety Act, 2015](#)



INDIA

For our most recent analysis of India’s legal framework, please see the dedicated section [p.47](#). Tech Against Terrorism had previously analysed India’s legal framework in [the Online Regulation Series 1.0](#).

India Legislative framework overview – Insights from previous ORS editions
Legal framework
<p>Amendments to the Guidelines for the Intermediaries and Digital media Ethics Code Rules, 2021</p> <ul style="list-style-type: none"> • 72-hour removal deadline for violating content, 15-day removal deadline for other complaints
<p>Guidelines for the Intermediaries and Digital Media Ethics Code Rules, 2021</p> <ul style="list-style-type: none"> • 36-hour removal deadline upon request from Indian authorities • Appointment of a Grievance Officer, Chief Compliance Officer, point of contact, and a Resident Grievance Officer • Easily accessible policies, updates to terms and policies • Publish compliance reports every 6 months • Establish an office in India • Notify users of content removal and explain why content was removed, have a redress mechanism • Messaging services are required to enable tracing of the original sender of the message • 72-hour deadline to provide assistance to authorised government agencies conducting investigations upon request
<p>Shreya Singhal v. Union of India, 2021</p>
<p>Information Technology Act, passed in 2000 and amended in 2005</p> <ul style="list-style-type: none"> • Exemption from liability for user-generated content, if platforms comply with government takedown guidelines



INDONESIA

For our most recent analysis of Indonesia’s legal framework, please see the dedicated section [p.74](#). Tech Against Terrorism had previously analysed Indonesia’s legal framework in [the Online Regulation Series 1.0](#).

Indonesia Legislative framework overview – Insights from previous ORS editions

Legal framework

Ministerial Regulation 5, 2020

- Legal liability for user-generated content if platforms do not comply with the law or cooperate with authorities
- Register with Kominfo and obtain an ID certification in order to operate in Indonesia
- Remove prohibited information or documents, or anything that could inform or provide access to prohibited content, and ensure that the service neither contains nor facilitates the dissemination of prohibited content
- 24-hour deadline to respond removal requests, or 4 hours for “urgent” requests (including terrorism-related requests). Failure to comply can result in fines and sanctions, including blocking of a platform’s services
- Appoint a local point of contact for content removal or data access orders
- Provide law enforcement and wider government access to electronic systems and data

Law No. 11 on Electronic Information and Transaction, 2008

Law No. 19 of 2016, amendment to the Law No. 11 of 2008

- Removal and blocking of content requests from Indonesian authorities
- Retention period for electronic content

Law No. 15 of 2003



UNITED STATES

For our most recent analysis of the US legal framework, please see the dedicated section [p.68](#). Tech Against Terrorism had previously analysed the US legal framework in [the Online Regulation Series 1.0](#).

United States Legislative framework overview – Insights from previous ORS editions

Legal framework

[AB 587, 2022](#)

- Bi-annual transparency requirements

[Texas House Bill 20, 2021](#)

- Restrictions on censoring a user and a user's speech
- Publish information about algorithms for displaying content
- Publish acceptable use policy
- Inform users of explanations of decisions to remove content
- User appeals

[Ohio House Bill 441, 2021](#)

- Restrictions on large social media companies “censoring” their users (including blocking and banning users based on what they post).

[Florida Senate Bill 7072, 2021](#)

[Updated Platform Accountability and Consumer Transparency Act, 2021-2022](#) [Platform Accountability and Consumer Transparency Act, 2020](#)

- Representative to take complaints through a phone number
- Publication of a quarterly transparency report

[Section 230 of the Communication Decency Act of 1996](#)

- Platforms are largely shielded from liability for user-generated content

[First Amendment](#) under the US Constitution



AUSTRIA

Austria is a new addition to the Online Regulation Series, you can find more detailed information about its legislative framework in the dedicated entry, [p.37](#).

Austria Legislative framework overview

Legal framework

Communication Platforms Act (Kommunikationsplattformen-Gesetz or KoPI-G), 2021

- 24-hour removal deadline for “obvious” illegal content, 7-day removal deadline for content which requires assessment
- Annual transparency reporting
- Appointment of a responsible officer and authorised recipient
- User complaints and redress procedures
- Financial sanctions for failure to comply

The Anti-Terrorism Package (Anti-Terror Packet), 2020

Reporting Office for Extremism and Terrorism, Directorate for State Security and Intelligence (DSN), 2015

The Symbols Act (Symbol Gesetz), 2015

Revisions to the National Security Strategy (Österreichische Sicherheitsstrategie or ÖSS), 2013



UNITED ARAB EMIRATES (UAE)

The UAE are a new addition to the Online Regulation Series, you can find more detailed information about its legislative framework in the dedicated entry, [p.61](#).

United Arab Emirates Legislative framework overview	
Legal framework	
Federal Decree Law No. 34 of 2021 on Combatting Rumours and Cybercrimes	
Internet Access Management (IAM) Regulatory Policy, 2017	
Federal Decree Law No. 2 of 2015 on Combatting Discrimination and Hatred	
Federal Law No. (7) of 2014	
Federal Law Number 5 of 2012 on Combatting Cybercrimes	
Federal Law No. (2) of 2006 on The Prevention of Information Technology Crimes	
Key legislative requirements for tech companies	
Access to platforms' services, or content on their services, may be blocked for users in the UAE if a platform is found to be in violation of UAE law	





SECTION 2.

COUNTRY ANALYSES

In this section we provide a deep-dive into the online regulatory frameworks and legal requirements for tech companies for all the jurisdictions analysed in this third edition of the Online Regulation Series. If a jurisdiction was analysed in a previous edition of the Online Regulation Series, the entry is marked as an update.



AUSTRALIA – UPDATE

Australia’s Online Safety Act came into effect in January 2022. We analysed the proposed legislation in our first and second Online Regulation Series, which you can see in our dedicated analysis [here](#). A further series of private members’ bills seeking to regulate social media companies, The Social Media (Basic Expectation and Deformation) Bill, the Social Media (Anti-Trolling) Bill and Social Media (Protecting Australian’s from Censorship) Bill, have failed to proceed.

Australia’s regulatory framework:

- **The [Online Safety Act](#)** came into effect in January 2022 and sets out to reform and expand existing online safety regulations. It does so by introducing five schemes to deal with different types of harmful online material, and by proposing the creation of a set of ‘online safety codes’.⁷ In September 2022, a group of six industry associations published the draft codes for public consultation, following the conclusion of which the industry bodies will use the feedback provided to amend the codes under consideration. The codes will then be submitted, for registration to Australia’s independent regulator for online safety, the Office of the eSafety Commissioner for registration. See the following section for a more detailed breakdown of the codes.
- Three private members’ bills seeking to apply more regulation to social media companies were presented between October 2021 and February 2022. All three proposals have lapsed following the dissolution of the 46th Parliament of Australia on 11 April 2022.
 - **The Social Media (Basic Expectations and Defamation) Bill** proposed that social media companies should be liable as publishers if they fail to take down allegedly defamatory material within 48 hours of receiving a notice from the eSafety Commissioner.
 - **The Social Media (Anti-Trolling) Bill** sought to allow courts to request users’ personal information, such as contact details and location data, from social media service providers if a user posts “potentially defamatory content”.
 - **The Social Media (Protecting Australians from censorship) Bill** proposed that foreign social media services should be prohibited from de-platforming content produced by members of the Parliament of Australia, political parties, journalists and election candidates. The Bill also proposed that these companies should be prohibited from censoring “philosophical discourse” on their platforms.

⁷ See our previous analysis (2021) for a detailed breakdown and analysis of the Act: [The Online Regulation Series 2.0 – Australia Update](#).

Key takeaways for tech companies

The incoming Online Safety codes outline minimum standards for tech companies to ensure users are protected from harmful content on their platforms. They cover eight sectors within the online industry, namely social media services, relevant electronic services, internet search engine providers, app distribution services, hosting services, internet carriage services and providers of computing equipment.

- *Addressing harmful content:* The codes are focused on the setting of minimum and optional compliance procedures for addressing Class 1 content, which refers to material which is deemed most harmful in accordance with Australia’s Classification (Publications, Films and Computer Games) Act 1995.⁸ If minimum compliance measures are not properly adhered to, service providers are liable to financial penalties.
- *Class 1 content:* The codes divide Class 1 content in two subcategories and outline categories of risk – Tier 1 being the most at risk of hosting Class 1 content.
 - o Class 1A refers to Child Sexual Abuse Material (CSAM), terrorist content and content which promotes extreme violence
 - o Class 1B content refers to material which depicts, promotes, or incites crime, violence, and drug-abuse.
- *24h removal deadline:* The most stringent compliance measures concern the proactive detection of Class 1A material on social media, message services and websites. Providers of such services are required to detect and remove Class 1A content within a 24-hour window, but for terrorist content the codes do allow some leeway, stating that material should be removed ‘as soon as reasonably practicable’.⁹

Tech Against Terrorism’s Commentary and Analysis

Freedom of Expression

The Online Safety Act mandates a 24-hour removal deadline for terrorist content. However, in the draft Online Safety Codes, industry bodies have included some acknowledgment of the time it might take to moderate and verify terrorist content if Class 1A content is to be removed ‘as soon as reasonably practicable’. At Tech Against Terrorism, we advise against the inclusion of short removal deadlines for terrorist content, which put significant pressures on small companies with insufficient technical capability and resources to meet them. Short removal deadlines also encourage increased reliance on automated removal of terrorist content, which can have a negative impact on freedom of expression. As false positives are highly likely with automated content moderation, companies require a layer of human moderation to ensure user content is not falsely removed.

The lack of explicit definition for terrorist content also poses risks to risks for freedom of expression. Neither the Online Safety Act nor the new industry codes for online safety provide a definition for tech companies to operationalise the removal of content, despite the codes calling for social media services, relevant electronic services, and designated internet services with larger user bases to remove this content as soon as possible. Without strict definitional boundaries around what can and cannot exist on their respective platforms, companies risk censoring their users by erring on the side of over-removal to comply with the legislation.

⁸ Tech Against Terrorism (2021), [The Online Regulation Series 2.0 – Australia Update](#)

⁹ Online Safety Australia, [Schedule 1 – Social Media Services Online Safety Code \(Class 1A and Class 1B Material\)](#).



Small Tech Companies and Proportionality

Whilst there has been some acknowledgment of proportionality in regulatory measures across the ecosystem of online service providers, existing Australian regulation does not provide enough clarity around this. For instance, the draft online safety codes do not fully consider the impact of minimum compliance measures on small companies with limited resources relative to larger companies. As enforcement of the codes is governed by financial realities, there is a chance that smaller companies could be fined for not having the resources to remove terrorist content within the necessary deadline or for not committing enough resources to maintaining a dedicated trust and safety function. As an organisation that works closely with smaller tech companies, we know that many of these companies are willing to address terrorist and violent exploitation of their platform but lack the requisite capacity. Australian regulation should ensure that these companies are supported in their efforts to counter terrorism and violent extremism, so that they can integrate more effectively into Australia's online ecosystem without putting a substantial strain on their resources.



AUSTRIA

Austria prohibits terrorist content online in legislation concerned with online hate (Hass im Netz), and principally in the Communication Platforms Act (Kommunikationsplattformen-Gesetz or KoPI-G). The KoPI-G, which came into effect on 1 January 2021, is a framework for protecting internet users by prohibiting illegal content on online services.¹⁰ Its definition of “illegal content” includes material concerning a “terrorist organisation”, “instructions for committing a terrorist offence”, “incitement to commit terrorist offences and approval of terrorist offences” - ‘terrorist offence’ having been defined under Austria’s Criminal Code (StGB).¹¹

Austria’s regulatory framework

Austria’s counterterrorism framework is comprehensive and was recently heavily influenced by the 2020 terrorist attack in Vienna.¹²

- [Revisions to the National Security Strategy \(Österreichische Sicherheitsstrategie or ÖSS\)](#), July 2013: The revisions to the National Security Strategy emphasised international cooperation in counterterrorism and cybercrime.
- [The Symbols Act \(Symbol Gesetz\)](#), 1 January 2015: Introduced as an effort to combat Islamist terrorism, the Symbols Act bans the use of Islamic State and al-Qaeda symbols. The Act was amended in January 2019, in December 2020 as part of the Anti-Terrorism Package, and again in July 2021.
 - The amendments broadened the Act’s scope to include symbols of Islamist association and those of groups such as the Muslim Brotherhood, Hamas, and Hezbollah, as well as the far right Identitarian Movement Austria. This latter amendment is notable for its discernible expansion of the scope of prohibited symbols to include those of far-right terrorist groups (see the full list [here](#)).
 - The Act prohibits people from displaying, wearing, or distributing symbols of such groups in public, including with the aid of electronic means of communication. Individuals posting violative content can receive fines of up to 4,000 EUR or one month of imprisonment.¹³
- [Reporting Office for Extremism and Terrorism, Directorate for State Security and Intelligence \(DSN\)](#), March 2015: The Federal Ministry of the Interior launched an initiative that allows online users to report “extremist and radical videos that have a connection to Austria” to the email address stopextremists@dsn.gv.at. The DSN reviews reported videos, initiates investigations, and flags reported videos to the operators.¹⁴ The DSN encourages users to report neo-Nazi, racist, and anti-Semitic content to the NS reporting office, at MAILTO: ns-meldestelle@dsn.gv.at.¹⁵

¹⁰ The law [defines](#) communication platforms under its scope as “an information society service where the main purpose or essential function is to enable the exchange of communications or performances containing ideas, whether spoken, written, audio or visual, by means of mass dissemination, between users and a wider range of other users”.

¹¹ Austria (2012), [Section 278f of the Criminal Code Instructions for committing a terrorist offense](#).

Austria (2012), [Section 278b of the Criminal Code Terrorist Association](#).

Austria (2012), [Section 282a of the Criminal Code, Incitement to commit terrorist offenses and approval of terrorist offenses](#).

¹² BBC (2020), [Vienna shooting: What we know about ‘Islamist terror’ attack](#).

¹³ Austria (2022), [Federal law prohibiting the use of symbols of the Islamic State group and other groups, Version 12/02/2022](#).

¹⁴ Bundesministerium Inneres, [Reporting Office for Extremism and Terrorism](#).

¹⁵ Bundesministerium Inneres, [Registration office – NS – re-activation](#).

- **The Anti-Terrorism Package (Anti-Terror Packet)**, 16 December 2020: The Package introduced new security measures following the IS-inspired terror attack in Vienna on 2 November 2020, which killed 4 and wounded 22. The Package broadened the Symbols Act and created a new criminal offence of “religiously motivated extremist connection”, targeting organisations that seek to overturn and replace the democratic constitutional order with a theocratic society and state. Participation in such an organisation is criminalised under Section 247b in the Criminal Code.¹⁶
- **The KoPI-G**: The Act was first announced in September 2020. The KoPI-G is part of a package of Austrian laws intended to tackle online hate (Hass im Netz). It came into effect on 1 January 2021 and gave social media platform operators until the end of March 2021 to implement its measures.
 - The KoPI-G is a framework for regulating illegal content on communication platforms, and includes in its definition of illegal content “[material relating to a] terrorist organisation, instructions for committing a terrorist offence, incitement to commit terrorist offences, and approval of terrorist offences”, as well as “hate speech”.
 - The Act includes other requirements for communication platforms, described in further detail under ‘Key Takeaways’ below.
- As a member state of the EU, Austria also complies with:
 - [EU Regulation on Terrorist Content Online 2021/784 \(TCO\)](#), June 2022.
 - [EU Digital Services Act \(DSA\), 2022](#): Due to the primacy of EU law, the DSA will supersede the KoPI-G.¹⁷

Main Regulatory Bodies

- [Communications Authority Austria \(KommAustria\)](#): The independent regulatory and ‘supervisory authority’ for the KoPI-G.
- [The Broadcasting and Telecom Regulatory-GMBH \(RTR\)](#): Founded on 1 April 2001 and “wholly owned by the federal state”.¹⁸ It is responsible for the KoPI-G complaints office, in addition to other media regulatory responsibilities.
- [Directorate for State Security and Intelligence \(Die Direktion Staatsschutz und Nachrichtendienst or DSN\)](#): Created in December 2021, by the abolition of the Federal Office for the Protection of the Constitution and Counter-Terrorism,¹⁹ the DSN is tasked with protecting Austria and its citizens from all forms of extremism and terrorism.

¹⁶ Counter Extremism Project (2022), [Austria: Extremism and Terrorism](#).

¹⁷ EUR-Lex, [Primacy of EU law](#).

¹⁸ RTR, [The Organisation](#).

¹⁹ Austria’s Federal Office for the Protection of the Constitution and Counterterrorism (Bundesamt für Verfassungsschutz und Terrorisusbekämpfung, or BVT): Austria’s domestic intelligence agency. The BVT was primarily responsible for combatting extremism and terrorism in Austria, in partnership with Interpol, Europol, and the EU Joint Situation Centre prior to its dissolution in 2021.

KoPI-G – Key Takeaways for Tech Platforms

- Removal timeframes: Platforms must remove reported content that is ‘obviously illegal’ within 24 hours of receiving the report. The law considers ‘obviously illegal’ content as that which “illegality is already obvious to a layperson without further investigation”. When illegality can be assessed only “after detailed examination”, platforms must remove content immediately after that detailed examination, and no later than seven days after receipt of the report.
- Transparency reporting:
 - Platforms must publish an annual transparency report on the handling of reports of alleged illegal content. For platforms with over one million registered users, the report must be published on a semi-annual basis.
 - Platforms must publish the reports no later than one month after the end of the period covered, and the report must be easily accessible to platforms’ users.
 - The Act details what must be measured in the report, including: the number of reports of alleged illegal content; the number that led to the deletion or blocking of content; the number of pieces of content reviewed; organisational information such as technical equipment, and the training, supervision, and competence of the staff responsible for processing reports and reviews.
- Review of reported content: Platforms must ensure that decisions to block and delete reported content are reached by an “effective and transparent” process. Platforms must notify of their decision both the users who reported content and any user whose content is actioned.
- Appointment of a responsible officer and authorised recipient: Platforms must appoint a “responsible officer” who will ensure compliance with the law. The officer must be easily and directly accessible to the supervisory authority. Platforms must also appoint an “authorised representative” capable of being served with warrants and notices issued by the authorities.
- Complaint mechanisms: Platforms must ensure users are able to submit complaints about the inadequacy of the reporting and/or review procedure. If complaints are unresolved by the platform or disputed by the user, users can contact the complaints office, managed by the RTR. The supervisory authority will investigate platforms who receive more than five “justified” complaints within a month.
- Fines: Platforms face fines of up to EUR 1 million for the failure to appoint a responsible officer, and fines of up to EUR 10 million for failures to comply with transparency reporting guidelines, to assess and block or remove illegal content, or to provide an adequate review procedure.
 - The supervisory authority takes into account the “financial strength” of a platform when issuing fines. It establishes a platform’s “financial strength” by reference to metrics such as the number of registered users and any previous violations, the level of the platform’s negligence, and the means by which employees are instructed to comply with the law.



Tech Against Terrorism's Commentary

'Obvious illegality' versus requirement for detailed examination

Adjudicating on the illegality of content and correctly identifying what requires immediate removal or further examination is a challenging task for tech companies to undertake at scale, particularly for terrorist and violent extremist content. Terrorist and violent extremist actors will often attempt to circumvent online counterterrorism efforts by posting "sanitised" or "grey area" content which stays within the bounds of legality or exploits the blurred lines between illegal and legal content.²⁰

Where the 'obvious illegality' of content is doubtful, platforms may choose to err on the side of caution and over-remove content to avoid fines, ultimately risking the removal of legal speech. Such action could result in increasing numbers of user complaints, generating a vicious cycle within the KoPI-G of over-removal, complaints, and fines.

Given that TVE content, and other types of content regulated by the KoPI-G, can often be difficult to identify correctly, and the likely result of such ambiguity being over-removal, regulation should instead focus on clearly defining what constitutes illegal or terrorist content online. Alternatively, regulatory authorities could focus on creating codes of practice to support the correct identification of content.

Tight removal deadlines risk violating freedom of speech

Reviewing the French "Cyberhate" law, the French Constitutional Court ruled in 2020 that 24 hours was too short a time limit to conduct an accurate assessment of a given piece of content's illegality.²¹ By requiring tech platforms to review reports of illegal content within 24 hours, Austria follows the trend of online regulation that fails to consider that the reasonable adjudication of illegal content requires both time and expertise.

Tech Against Terrorism cautions that tight removal deadlines, in tandem with stiff penalties for companies who are unable to moderate their platforms, promote an overabundance of caution out of which platforms are inclined to over-remove content. This significantly infringes human rights, particularly freedom of expression, as legal online content may be removed.

To comply with tight removal deadlines, platforms may also over-rely on automated content moderation tools to undertake moderation at scale. Automated tools can often lack the nuance of human moderation, do not entertain considerations of local context and language, and should be paired with mechanisms of human review to mitigate the risk of infringement on freedom of expression.

²⁰ Terrorists and violent extremists are generally fully aware of platforms' content moderation rules and enforcement practises and will sometimes try to circumvent such rules by posting content within the limits of what is acceptable on a platform to avoid deplatforming. "Grey area" content can be a piece of content that is produced in support of a terrorist or violent extremist group, but which does not make this support explicit or call directly for violence. The term also refers to content that can be considered legal but "harmful".

²¹ To read more about this, please see our [entry on France](#) in the Online Regulation Series 2020.



Transparency Reporting

Tech Against Terrorism commends the KoPI-G’s commitment to promoting transparency and accountability. However, any mandate to publish transparency reports must be greeted with caution, since a ‘one-size-fits-all’ approach is not compatible with transparency. Policymakers should consider the practical challenge of collecting the necessary data if the policies and processes needed to underpin the production of such a report are not in place.

The KoPI-G’s transparency guidelines are also overly focused on quantitative metrics – for instance, the number of user reports, content removals, and account removals. Whilst simple reporting on numbers of content flagged and removed does provide a credible base for transparency reporting, Tech Against Terrorism recommends that tech companies provide contextual information to explain the metrics and policy framework to users when it is within the platform’s capacity to do so.²² This additional information should be included in the transparency report in order to provide a layer of “meaningful transparency” to the practice of moderation on a given platform.²³

For more information on transparency reporting, please see Tech Against Terrorism’s [Transparency Reporting Guidelines](#).

Concerns over freedom of expression and state control of online speech

The KoPI-G makes no commitment to defending freedom of expression online. Alongside the threat of heavy fines and a focus on quantitative transparency data, this abstention could license the over-removal and blocking of content without regard for the risk of infringement on human rights. The Austrian government should ensure that the freedom of expression is not compromised by the implementation of the law.

The Act gives tech platforms great power to assess the’ illegality of content, whereas restrictions to freedom of expression ought, in accordance with international standards, to be determined by independent judicial bodies only. Outsourcing the adjudication of legality to private companies, rather than confining this jurisdiction to the legal system, risks undermining the rule of law. According to David Kaye, this lack of judicial oversight is incompatible with international human rights law.²⁴ Whilst Tech Against Terrorism commends the inclusion of a user appeals process, which is integral to maintaining respect for the freedom of expression in content moderation efforts, the final decision on appeals lies with the RTR, “wholly owned by the federal state”. It would be preferable for the power to determine the legality of content to be conferred on independent tribunals, by whom the difficult and sensitive decisions to interfere with fundamental rights are more properly made.²⁵

²² For more resources on transparency reporting, please see our [Transparency Reporting Guidelines](#).

²³ Svea Windwehr and Jillian C. York (2020), [Thank You For Your Transparency Report, Here’s Everything That’s Missing](#), The Electronic Frontier Foundation.

²⁴ David Kaye was the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression from August 2014 to July 2020. See more [here](#).

²⁵ David Kaye and Jason Pielemeier (2020), [The Right Way to Regulate Digital Harms](#), Project Syndicate.



EUROPEAN UNION – UPDATE

Since our first analysis of the EU regulatory framework in 2020, regulation in the EU has evolved from ambitious proposals to become an extensive and structured reality which will profoundly impact the practice of content governance both in the EU and beyond. This new regulatory framework is chiefly based on the Regulation on Terrorist Content Online (also known as TCO), passed in 2021, and the Digital Services Act (DSA), passed in 2022.

This entry focuses on providing key updates on the final versions of the EU's recent online regulation and the state of their implementation by the EU and its Member States as of December 2022.

The regulation on addressing the dissemination of terrorist content online

The EU's [Regulation 2021/784 on addressing the dissemination of terrorist content online](#) (often called the "TCO" regulation) was enacted in June of 2021, and came into force on 7 June 2022. The principal provisions of the TCO require services, which qualify as hosting service providers (HSPs), to:

- Act on removal orders from competent national authorities within one hour of receipt.
- Introduce "specific measures" to prevent terrorist content if instructed to do so by competent authorities.
- Preserve removed terrorist content for six months.
- Produce transparency reports on measures taken to comply with the regulation.
- Inform users when their content has been removed in the name of compliance with the regulation, and introduce complaint mechanisms for users whose content has been removed.
- Establish a point of contact to coordinate and respond to removal orders from competent authorities, as well as assign a legal representative in the EU.
- Inform national authorities promptly when becoming aware of terrorist content involving an imminent threat to life situation.

A central provision of the TCO is the empowerment of national competent authorities in the EU to send HSPs removal orders for terrorist content. On the day the TCO came into force, the EU published the list of said national competent authorities, as provided by the EU Member States. As of December 2022, competent authorities for some of the Member States are yet to be added to the list, though the EU Commission notes that this list is to be regularly updated upon the reception of notifications from EU Member States.

Tech Against Terrorism Europe

In October 2021, the EU Commission announced funding for initiatives aimed at supporting smaller platforms to comply with the TCO. Tech Against Terrorism welcomed this announcement given our longstanding concerns about the risks of smaller platforms being penalised for their lack of resources and understanding of the threat landscape. Tech Against Terrorism is pleased to announce that we have been selected to provide this support with the launch of Tech [Against Terrorism Europe](#) (TATE). Via TATE, a consortium with partners across Europe, Tech Against Terrorism will provide capacity-building support for small and micro platforms to align their online counterterrorism response with the European Union's TCO requirements. This will be modelled on our [existing mentorship programme](#) through which we support platforms in strengthening their online counterterrorism framework in a manner grounded in transparency and the protection of fundamental rights.



The Digital Services Act

Building on the approaches developed in the 2001's eCommerce Directive, the DSA provides a comprehensive set of new rules for all digital services, including social media, online marketplaces, and other online platforms operating in the EU. The DSA is particularly concerned with content governance and the impact of the online sphere on democratic processes. As such the DSA outlines a number of requirements for the moderation of content on digital services and the mitigation of risks to their users; in the words of the EU Commission's President, Ursula von der Leyen, the DSA "gives practical effect to the principle that what is illegal offline should be illegal online".

Following nearly two years of negotiations, the final version of the Digital Services Act was approved by [the EU Parliament on 5 July 2022](#). The DSA was later approved by the EU Council [and published in the EU Official Journal](#) on 27 October 2022, and came into force on 16 November 2022. From this date, platforms operating in the EU have three months to report the number of active users on their services for the EU Commission to assess which platforms are to be designated as very large platforms or search engines (respectively VLOPs and VLOSEs). Once the designation process has been completed by the EU Commission, VLOPs and VLOSEs will have four months to comply with the DSA requirements. The DSA is to fully come into force and apply to all service providers in the EU on 17 February 2024. EU Member States are to nominate their Digital Services Coordinator by this date.²⁶

Key takeaways – DSA

To support tech companies in understanding how the DSA will impact their response to terrorist and illegal content, the below section provides an overview of the main provisions relevant to content moderation. It also focuses on provisions added since the Online Regulation Series 2.0, published in 2021.

- **Parity of illegality:** The DSA upholds the parity of illegality offline and online, defining illegal content as content that "in itself or in relation to an activity [...] is not in compliance with Union law or the law of any Member State". This means that what constitutes illegal content is defined in national or EU laws, including with the example of terrorist content.
- **Removal without "undue" delay:** Platforms must action removal orders for illegal content from Member States without undue delay. Whilst there is no specification as to a timeline for removal, which is heavily reliant on the situation and type of content, the DSA does refer to the 24h removal timeline recommended by the 2016 Code of Conduct on Countering Illegal Hate Speech Online.
- **No general monitoring obligation, nor general legal liability provisions:** Whilst tech companies are required to act against illegal content when made aware of it, including via removal requests from Member States, Article 8 excludes any general monitoring or active fact-finding obligation.²⁷ Platforms are also protected from legal liability for content shared on their services as long as they do not interfere with the content and its transmission, and do not have knowledge of illegal content and apparent illegal activity.

²⁶ European Commission (2022), [Digital Services Act: EU's landmark rules for online platforms enter into force](#)

²⁷ The DSA specifies that knowledge or awareness of illegal content cannot be considered in the general sense of awareness of illegal content on the service, but rather is limited to specific uses in cases against which the service provider has not acted.

- **Community guidelines requirements:** The DSA does not mandate platforms to take action on harmful content beyond what is illegal in the EU and its Member States. However, Art. 14 enables platforms to action content based on their own Content Standards and includes provisions which stipulate that Content Standards must be clear and accessible to users:
 - Platforms should publish information about policies and enforcement procedures, including the measures and tools used for content moderation and review.
 - Platforms must also act on Content Standards, in a diligent and proportionate manner that accounts for the fundamental rights of the users, as per the EU Charter.
- **Transparency reporting:** Articles 15 (intermediary services), 24 (online platforms), and 42 (VLOPs) outline transparency reporting obligations:
 - Platforms are to publish yearly transparency reports on their content moderation actions, the removal orders received and actioned, user complaints received and actions taken, as well as information about the use of automated tools, including removal errors rate and relevant safeguards.
 - Art. 24 states that the EU Commission may publish templates for transparency reporting.
- **Requirements for VLOPs and VLOS:** The DSA includes specific requirements for platforms and search engines with more than 45 million average monthly active users in the EU. Amongst other requirements, VLOPs will have to:
 - Conduct assessments of any systemic risks posted to the EU, with a particular view to understanding how their services may impact civic discourse and fundamental rights, as well as how illegal content is disseminated on their services.
 - Act on the risk assessments by implementing risk mitigation strategies where required, including at the product design and content moderation level.
 - Assess how their service can be used to contribute to a significant threat to the EU when a crisis response mechanism is triggered and, accordingly, implement measures to limit and eliminate the threat.
 - Conduct a yearly independent audit.
 - Make available a user complaint mechanism for a period of at least 6 months.
- **Crisis response mechanism for VLOPs (Art. 36).** Where a crisis occurs, the Commission, acting on a recommendation of the Board of national authorities, can require one or more VLOPs to take specific actions to address the threat.
 - The DSA defines a crisis as “extraordinary circumstances lead[ing] to a serious threat to public security or public health in the Union or in significant parts of it” (Art. 36.2). This can include acts of terrorism or emerging acts of terrorism.
 - The request to initiate a crisis response will specify a period within which specific measures referred are to be taken, not exceeding three months. Crisis response mechanisms can be renewed.
 - This mechanism can require platforms to remove war propaganda, for instance.²⁸ However, the choice of the specific measures to be taken to limit the impact of a crisis remains with the service providers. The role of the Commission is to oversee and assess the effectiveness of the measures taken.

²⁸ In practice, this crisis mechanisms is similar to the EU imposed sanction on Russian state-owned outlets Russia Today and Sputnik. See: Council of the European Union (2022), [EU imposes sanctions on state-owned outlets RT/Russia Today and Sputnik's broadcasting in the EU](#)

- **Out of court dispute settlement:** When a user has exhausted all internal complaint avenues with a VLOP, it may seek an out-of-court dispute settlement free of charge – the platform will have to pay the fee if in the wrong. National Digital Services Coordinators, as established by the DSA, will certify out-of-court bodies in the countries where they are based. Small and micro-tech platforms will be exempted from this provision.
- **Single points of contact:** Platforms are to designate a single point of contact to enable communication with Member State authorities. Platforms must also designate a single point of contact for users to communicate with them. In both instances, information about the points of contact must be public and easily accessible.

Tech Against Terrorism's Commentary

What is illegal offline should be illegal online

The DSA commendably confines moderation and removal requirements to illegal content, and refrains from demanding platforms to act against content that is not strictly illegal under Member State or EU law. Tech Against Terrorism has long advocated for online regulation to safeguard the rule of law and due process by not mandating platforms to remove content when there is no clear legal basis to do so, and by refraining from making what is legal offline illegal online. Tech Against Terrorism also commends the explicit lack of requirements for systematic proactive monitoring of online content, as well as the DSA's removal requirements being based on Member States' judicial and administrative authorities' orders rather than on platforms' own adjudication of content illegality. The DSA sets a clear legal framework for mandatory content removal by tech companies based on the rule of law, as well as clear expectations for users as to what content platforms are legally required to remove.

The DSA also enables service providers to set their own rules for moderation in accordance with their Content Standards, subject to the requirement to respect fundamental rights. This is positive, as it allows platforms to act rapidly against harmful content they may discover on their services. However, the question of "grey area content" remains unanswered and at risk of exploitation by terrorists and violent extremists. It will be interesting to see if this prompts more regulation and voluntary codes of practices around certain types of grey area or borderline content in the future, as is done currently with the Code of Practice on Disinformation which is envisioned by EU regulators as an essential complement to the DSA.

Whilst Tech Against Terrorism generally commends the DSA framework for illegal content removal, we are concerned with the risks of abuse by Member States attempting to exploit it as a tool for extra-jurisdictional application of national law. Art. 9, concerning orders to remove illegal content, specifies that national authorities are to include an explanation as to why content is illegal under national law or EU law and that the territorial scope of the order is to be limited to what is strictly necessary to achieve the objective of any such order. However, there is a risk that platforms are pressured to comply with removal orders even when they do not strictly meet the conditions referred to in Art. 9. There is also a risk that platforms, due to a lack of resources and capacity to review the validity of the orders, act on all orders by default to avoid penalties and err on the side of over-removal with a negative impact on fundamental rights. The DSA does provide an effective right of redress, to avoid both under- and over-removal of content on grounds of illegality. Tech Against Terrorism further recommends the EU Commission to ensure a strict process to regularly review removal orders from Member States to mitigate against abuse.



A coordinated crisis response mechanism

The Russian invasion of Ukraine prompted EU policymakers to add a crisis response mechanism to the DSA, whereby very large online platforms are required to mitigate threats to public security or public health in the EU, including acts of terrorism and “emerging” acts of terrorism.

By including terrorism as a potential trigger for the crisis response mechanism, the DSA builds on the voluntary EU crisis protocol for responding to terrorist content online²⁹ adopted in October 2019. However, the DSA crisis mechanism is likely to have a more significant impact on the EU online space as all VLOPs will be required to act on it and because the default period will be 3 months (the mechanism being renewable). The first introduction of the crisis response mechanisms in the DSA in April 2022 drew strong criticism from civil society concerning the impact on the rights to freedom of expression and access to information, as well as on the rule of law. Whilst the final version of the DSA appears to have addressed the concerns raised,³⁰ careful consideration should be brought to the review of future implementation of the crisis protocol.

Tech Against Terrorism calls for the crisis response mechanism to be further detailed with regard to its application to terrorism, and specify precisely when a terrorist act or emerging terrorist act capable of triggering the crisis mechanism takes place. The threshold for an emerging terrorist act must be clarified to prevent a continuous state of crisis justified by anticipated terrorist concerns.

Transparency reporting

Tech Against Terrorism welcomes the emphasis given to clear policy and transparent enforcement in the DSA. However, we caution against the stipulation of an inflexible reporting template. Transparency reporting should acknowledge and reflect the tech sector’s diversity, both in product offering and in approaches to content moderation; there cannot be a universally applicable solution to transparency reporting, since such a solution would force companies to squeeze their data into predetermined categories which will render transparency reporting meaningless. The DSA already includes detailed guidelines as to the type of information and metrics platforms should report on; the EU Commission should confine its requirements to the basics and allow the tech sector to reflect its diversity in its transparency reporting.

Tech Against Terrorism also calls on governments to heed their calls for transparency reporting and to report on their online counterterrorism efforts in line with our [Transparency Reporting Guidelines](#). EU policymakers should in particular encourage DSA national coordinators to report on their country’s removal orders.

²⁹ European Internet Forum (2023), [The Revised EU Crisis Protocol: Responding To Terrorist Content Online](#).

³⁰ Allen Asha (2022), [The Digital Services Act: Political Agreement Reached, Long Road Ahead Awaits](#), Center for Democracy & Technology



INDIA – UPDATE

Since we published our first analysis of India’s regulatory framework in the Online Regulation Handbook (July 2021), India has presented [Amendments to the Information Technology Rules \(IT Rules\) 2021](#) and published a [Draft Telecommunication Bill](#). Additionally, the Indian government is developing the Digital India Act (DIA) to replace the [Information Technology Act 2000](#) as the primary source of regulation for the entire digital ecosystem, and a revised Digital Data Protection Law.

India’s Regulatory Framework

- **The [IT Rules \(Intermediary Guidelines and Digital Media Ethics Code\)](#)**, February 2021, create a new regulatory framework for online content. These rules formalise what online content is prohibited in the country and allow Indian authorities to request content removal. Tech companies were given three months, dating from the law’s enactment on 25 May 2021, to comply with the 2021 Guidelines.
 - In July 2022, the [Ministry of Electronics and Information Technology](#) released [draft amendments](#) to the IT Rules for consultation.
 - On 28 October 2022, the government [presented](#) amendments to the IT Rules 2021, which came into force immediately.
- **The [Draft Telecommunication Bill](#)** proposed by the Indian government seeks to modernise regulation of the telecommunication sector by consolidating and amending the laws “governing provision, development, expansion and operation of telecommunications, services, networks and infrastructure.
 - Telecommunication services are those “made available to users by telecommunication” and include conventional phone and SMS services, as well as “over-the-top (OTT) communications services” including WhatsApp, Signal and Facetime.
 - The Bill grants the government expansive powers including that of directing the interception and disclosure of any messages “on the occurrence of any emergency or in the interest of public safety.” The Bill also grants the government the right to shut down the internet through suspending telecommunication services if the government “is satisfied that it is necessary or expedient to do so, in the interest of national security, friendly relations with foreign states, or in the event of war.”
 - The Indian government [sought](#) feedback on the draft Bill from the general public, various stakeholders and industry associations.
- The Ministry of Electronics and IT (MeitY) is currently working on a digital regulatory framework, the Digital India Act (DIA), a draft of which is expected in early 2023. This will replace the [Information Technology \(IT\) Act 2000](#) and cover the entire digital ecosystem, from social media platforms, OTT platforms, and online apps, to the metaverse and blockchain-based technologies. The proposed Act will reportedly target cyber offences including cyberterrorism, as well as misinformation and the incitement of violence.

³¹ Tech Against Terrorism (2021), The Online Regulation Series Handbook, p.78

³² Kashyap Hemant (2022), [Digital India Act Will Monitor Social Media, Metaverse, OTT Platforms: Report](#), Inc42



Key takeaways for tech companies:

Amendments to the Information Technology Rules (IT Rules) 2021

- **The Grievance Appellate Committee (GAC).** This mechanism is designed as an avenue of appeal against the decisions made by platforms' grievance officers on user reports of violating content.³³
 - The Central Government is required to establish one or more Committees within three months of the amendments, with each consisting of a chairperson and two members appointed by the government. One of these members must be ex-officio and two must be independent.
 - When a complainant is dissatisfied with a decision by a grievance officer, they can take the case to a Committee, who should “endeavour to resolve the appeal” within 30 days.
 - The concerned tech company must comply with the order passed by the Committee with no appeal possible.
- **A shortened removal deadline for violating content:** The Ministry has established a stricter timeline for tech platforms to remove prohibited content expeditiously.
 - Complaints containing a removal request must be addressed within 72 hours.
 - All other complaints must be acknowledged within 24 hours and resolved in 15 days.
- **Expanded obligations for intermediaries to ensure enforcement of 2021 Rules:**
 - The 2021 Rules require intermediaries to “inform” users about specific restrictions (as defined in the 2021 Rules) about the types of content they are allowed to create, upload, or share. The amendment states intermediaries should do this “periodically, and at least once in a year.”³⁴
 - The amendment will expand the obligation on intermediaries to “make reasonable efforts to cause the user... not to host, display, upload, modify, publish, transmit, store, update, or share any [prohibited] information”. This includes (but is not limited to) content that is:
 - “Obscene, pornographic, paedophilic, invasive of another’s privacy including bodily privacy, insulting or harassing on the basis of gender, racially or ethnically objectionable, relating or encouraging money laundering or gambling, or promoting enmity between different groups on the grounds of religion or caste with the intent to incite violence;
 - Harmful to child [sic]
 - Threatening to the unity, integrity, defence, security or sovereignty of India, friendly relations with foreign States, or public order, or causes incitement to the commission of any cognisable offence, or prevents investigation of any offence, or is insulting to other nations;
 - Violates any law for the time being in force;”³⁵
 - Additionally, intermediaries are required to “respect the rights guaranteed to users under the Constitution of India” and take all reasonable measures to ensure accessibility of their services to all users along with “due diligence, privacy and transparency.” This includes making their privacy policy and user agreements available in the 22 Indian languages covered in the “Eighth Schedule”.³⁶

³³The IT Rules require tech companies to appoint grievance officers as representatives for addressing reported violations of the Rules.

³⁴Government of India, <https://egazette.nic.in/WriteReadData/2022/239919.pdf>

³⁵Government of India, <https://egazette.nic.in/WriteReadData/2022/239919.pdf>, Rule 3(a)(i)(b)

³⁶Government of India, https://www.mha.gov.in/sites/default/files/EighthSchedule_19052017.pdf



Draft Telecommunication Bill, 2022

- In situations involving public emergencies, public safety, or national security concerns, the Bill authorises the time-limited suspension of transmission of messages, or of provision of telecommunication networks of services by the government.
- Under these situations, the government is authorised to “direct that any message or class of messages, to or from any person or class of persons, or relating to any particular subject... be intercepted or detained or disclosed to the officer mentioned in such order.”

Tech Against Terrorism’s Commentary:

Government control over online speech

Both tech platforms and human rights advocates have expressed concerns that the Grievance Appellate Committee (GAC) could confer on the government excessive control over online speech. A [joint letter](#) submitted to the Indian government by digital rights organisations argues that the provision empowers an unelected body that is not independent from political influence to control protected speech. The Committee, in its proposed form, lacks independent judicial oversight over demands to remove content or any remedy against its decisions. Governments should ensure that regulation complies with international norms around freedom of expression and has strong safeguards against abuse and censorship of legal and legitimate content online.

The Asia Internet Coalition - an industry group representing Apple, Meta, Google, Amazon, Twitter and Spotify - openly criticised the proposed amendments and urged the Indian government to allow tech companies to adopt their own self-regulatory grievance redressal system instead. In a letter to the government, they highlight a lack of clarity on the extent or limitations of the GAC’s powers, with the risk of it acting as “a judicial or quasi-judicial body with wide-ranging authority.” Tech Against Terrorism encourages the Indian government to engage in open consultation with digital rights advocates and civil society, for the particular purpose of clarifying limitations on the GAC’s powers and devising an appeal mechanism for intermediaries.

Due diligence obligations incentivise overzealous content removal

The due diligence obligations established in the IT Rules 2021 as a condition of platforms’ non-liability for user content have been expanded in the amendments. Civil society actors have construed the duty in the proposed amendments for intermediaries to ‘ensure compliance’ with rules, regulations and policies as a particular threat to freedom of speech. Although the wording in the final amendments has changed to impart a duty to ‘take reasonable efforts to prevent users from uploading such content’, this language still suggests intermediaries would be required to regulate online speech irrespective of specific complaints, encouraging proactive monitoring that could lead to removal of legitimate speech.⁴²

³⁷ Government of India, <https://dot.gov.in/sites/default/files/Explanatory%20Note%20to%20the%20draft%20Indian%20Telecommunication%20Bill%2C%202022.pdf> Ch. 6 no. 42, p 14.

³⁸ Government of India, <https://dot.gov.in/sites/default/files/Draft%20Indian%20Telecommunication%20Bill%2C%202022.pdf> clause 24, 2(a)

³⁹ Kaye David (2019) [Promotion and protection of the right to freedom of opinion and expression : note by the Secretary-General](#), UN. Human Rights Council. Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression

⁴⁰ Singh Manish (2022), [American internet giants seek changes to India’s proposed new IT rules](#), TechCrunch

⁴¹ Ibid.

⁴² Access Now (2022), [Civil society calls on Indian government to withdraw amendments to IT Rules](#)

Additionally, the categories of information users are required not to host are broad, including ‘racially and ethnically objectionable’ and ‘harmful to child’[sic], and vague, and therefore encourage the over-removal of speech online to avoid non-compliance. Tech Against Terrorism encourages governments to ensure that definitions are clear and narrow, and that provisions can be made operational by all platforms in scope of the regulation.

Short removal deadlines

The risk of overzealous content removal is exacerbated by the short timeframe proposed which states that “complaints relating to 3(1)(b) shall be acted upon expeditiously and redressed within 72 hours of reporting.” Short timeframes for the removal of content are a common yet unrealistic expectation being introduced in various jurisdictions, and which are especially problematic for smaller platforms lacking resources to effect compliance. Given the huge volume of content involved and nuanced response needed to adjudicate it, this provision will encourage over-censorship by platforms in order to comply with the regulation. It may also incentivise additional reliance on automated tools, such as algorithmic filtering, which can lead to the removal of legitimate speech and tends to discriminate heavily against regional languages.

Right to privacy

Tech companies and digital rights experts have already criticised the traceability requirement present in the 2021 IT Rules, which mandates tracing the originator of a message. Critics have, in particular, flagged the practical incompatibility of such proposals with end-to-end encryption (E2EE), given the burdensome technical changes needed to comply with this requirement for platforms that do not collect metadata. The Draft Telecommunication Bill has faced similar criticism from digital right experts, some of whom have called it “yet another attack on end-to-end encryption, and people’s fundamental rights and freedoms.” Specifically, the provision which authorises the government to direct the interception and disclosure of any messages could be used to undermine encryption and the right to privacy.

OTT services that offer E2EE such as WhatsApp and Signal would be forced to remove or circumvent this feature in order to comply with these rules, or alternatively exit the Indian market. WhatsApp has already filed a legal complaint contesting the 2021 Guidelines and specifically the traceability requirement arguing that it violates the privacy rights protected by the Indian Constitution.

Tech Against Terrorism is concerned that mandating companies to track encrypted messages risks compromising existing encryption protocols and traceability assurances. This presents risks to tech platforms and their users, whose security and privacy could be compromised.

You can find our landmark report on assessing terrorist use of end-to-end encrypted (E2EE) services – including our recommendations on mitigation strategies for tech companies and policymakers, as well as our review of global legislation impacting E2EE – [here](#).

⁴³ Ibid

⁴⁴ Ibid

⁴⁵ Access Now (2022), [India’s Draft Telecommunication Bill must be revamped to protect human rights](#)

⁴⁶ Menn Joseph (2021), [WhatsApp sues India govt. says new media rules mean end to privacy](#), Reuters (republished in the Financial Post)



KYRGYZSTAN

Kyrgyzstan's approach to countering terrorist and extremist content online is heavily focused on the removal of content and the blocking of access to websites and platforms. Kyrgyzstan's definition of terrorist and extremist activity is broad, thus risking the suppression of legitimate political speech and the curtailment of human rights such as freedom of expression. The definition of TVE is therefore reflective of a broader issue of definitional ambiguity in national regulatory frameworks.

Kyrgyzstan's regulatory framework

- **Law on Countering Extremist Activity 2005**, originally enacted in 2005, this legislation has been amended several times, most recently in 2016. It criminalises the public expression of support for extremist activity.
- **Law on Combatting Terrorism**, enacted in 2006 and amended in 2009. The law provides the legislative framework for Kyrgyzstan's approach to counterterrorism by defining terrorism and terrorist activity. In 2020, the government proposed a draft amendment to the law which sought to expand the scope of terrorist activity to include the support of any organisation whose activities the government deems terroristic along with a specific prohibition of 'terrorist materials'.⁴⁷
- **Criminal Code of Kyrgyzstan**,⁴⁸ last amended in 2017, the code further and more specifically criminalises 'hate speech', which is included under the definition of extremist activity in the 'Law on Combatting Terrorism'. Article 315 of the code criminalises the production, dissemination, transportation, shipment and purchase (with the intent to distribute) of extremist materials. This has implications for online service providers operating within its jurisdiction, as they are often used to host and disseminate TVE content and are considered complicit.
- **Law On Protection from False and Inaccurate Information, 2021**.⁴⁹ Whilst this law is focused on the blocking and removal of 'false or defamatory content' and not explicitly terrorist and extremist content, it nonetheless highlights Kyrgyzstan's current and future approach to regulating content online.

Relevant bodies

- **State Committee for National Security (GKNB/SCNS)**: Responsible for upholding national security through the prevention and investigation of terrorism and organised crime in Kyrgyzstan.
- **State Committee of Information Technologies and Communication (SCITC)**: Responsible for digitalisation and policy coordination across all digital infrastructure and information technology.
- **Ministry of Culture, Information, Sport and Youth Policy**: Responsible for preserving and promoting Kyrgyzstani culture, including through mass communications. Under the recent false information law, it has been given the authority to request content removal and to issue orders to block platforms if they fail to comply.⁵⁰

⁴⁷ Article 19 (2020), [Kyrgyzstan: Draft Law on Countering Terrorism](#)

⁴⁸ Article 19 (2021), [Kyrgyzstan: Report on freedom of expression and 'extremism'](#)

⁴⁹ Freedom House (2021), [Freedom on the Net 2021 – Kyrgyzstan](#)

⁵⁰ Freedom House (2022), [Freedom on the Net 2022 – Kyrgyzstan](#)



Key takeaways for Tech Platforms

- **Definitional ambiguity and blocking:** Tech platforms should be aware that Kyrgyzstan often uses blocking to stop platforms from hosting content, which is illegal or otherwise objectionable, the legal basis of which is contained in the law ‘on protection from false and inaccurate information’. In the context of terrorist content, Kyrgyzstan’s definitions of terrorism, extremism and extremist activity are overlapping and fundamentally broad in scope. This has implications for tech companies operating in the country, as the failure to remove content, however nebulously defined, can result in being blocked and restricted from providing services to users within the country.
- **24-hours content removal limit:** The law ‘on protection from false and inaccurate information’ allows users of a social media platform to request the 24-hour removal of a specific piece of content which they deem to be false. If platforms fail to comply, the user can request that the platform remove the account responsible for the specific piece of content. Additionally, the law allows for the blocking of whole platforms and sites for lack of compliance within the allotted time frame.

Tech Against Terrorism’s Commentary

Definitional ambiguity

Kyrgyzstan’s ‘Law on Countering Extremist Activity’ defines extremism as “planning and carrying out activities designed to change the basis of the constitutional system by force, to undermine security in a country and to seize and usurp power”.⁵¹ It defines extremist activity as an action that stokes “ideological, political, racial, national (ethnic) or religious hatred or enmity”. The extensive list included in this legislation includes ill-defined activities such as “carrying out terrorist activities”, “humiliating national dignity”, “acts of vandalism” and “hooliganism”.⁵² The breadth of this definition significantly widens the range of content that companies can be required to remove, and has extensive implications on users’ human rights, namely freedom of expression, access to information, political opinion and their right to assembly.

Website blocking and freedom of expression


Widespread content blocking is used as a tool by the Kyrgyzstan government to quell extremist activity and rhetoric. In many cases, this involves blocking access to entire websites and platforms through court orders. As reported by Freedom House, over the last 5 years, these court orders have mandated the blocking of several hundred websites and platforms, such as Soundcloud, Internet Archive, JustPast.it and Change.org.⁵³ Authorities have also threatened to block URLs leading to Twitter and YouTube, but have not followed this through.⁵⁴ Blocking users from accessing platforms in their entirety based on an overbroad and ambiguous definition of extremist activity has serious implications for human rights, in particular the rights to freedom of expression and access to information. Tech Against Terrorism recommends that governments follow a targeted and proportional approach to countering terrorist use of the internet rather than blocking online service providers in their entirety, which may lead to TVE users migrating to other services and thus will displace the threat rather than tackle it.

⁵¹ <https://legislationline.org/taxonomy/term/11615>

⁵² Human Rights Watch (2018), ‘We Live in Constant Fear’ - Possession of Extremist Material in Kyrgyzstan

⁵³ Freedom House (2022), [Freedom on the Net 2022 – Kyrgyzstan](#)

⁵⁴ Ibid



Enforcement of the law on protecting against false information suggests that the over-removal of content on social media platforms also poses a challenge to freedom of expression. At present, this explicitly applies only to content that is deemed ‘false’ or ‘inaccurate’ in accordance with the recent false information law. However, there is a high likelihood that companies having the capacity and resources to remove content capable of falling within the Kyrgyzstani definition of extremism could pose an equal risk to users’ freedom of expression.

Right to privacy

The new law on protecting against false and inaccurate information mandates that social media companies and online service providers send all users’ personally identifiable information to a single government-owned registry system, to give the government of Kyrgyzstan more oversight and power to prosecute individuals for their online activity. By instituting such a requirement, the government has undermined the privacy and security of users’ personal data. This registry is highly vulnerable to compromise by malicious cyber actors and could be used by the government to silence political opponents and citizens with opposing views.

NEW ZEALAND – UPDATE

New Zealand is currently in the process of developing a new, holistic regulatory framework for content moderation, with the aim of tackling harmful content across all media from print and broadcast to digital platforms. The new proposed framework, further detail of which is yet to be announced, is likely to combine different approaches “spanning Government, co-regulatory and self-regulatory approaches.”⁵⁵ It may also incorporate the recently published [Aotearoa New Zealand Code of Practice for Online Safety and Harms](#), a self-regulatory voluntary code developed by leading tech companies (including Meta, Twitter, and YouTube) in cooperation with the New Zealand government.

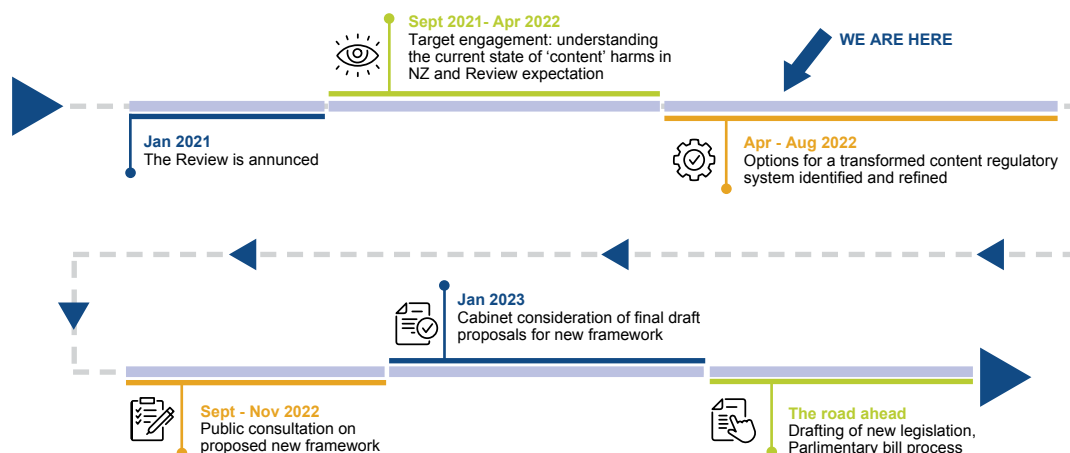
This entry addresses changes to New Zealand’s regulatory framework implemented since publication of the Online Regulation Series 2.0 in November 2021. For a complete overview of New Zealand’s regulatory framework please see the dedicated entry in the Online Regulation Series 2.0.⁵⁶

The Content Regulatory Review

In June 2021, New Zealand’s Department of Internal Affairs (DIA) announced a Content Regulatory Review, “to design and implement a new approach to content regulation that minimises the risk of harms caused by content to New Zealanders.” This new approach endeavours to simplify New Zealand’s complex ‘mixed-model’ approach to content regulation, where different ‘rules’ apply for content on different platforms. Current regulations are designed around traditional ideas of content (books, magazines, TV etc.), and lack the flexibility to respond to the various forms of online content. By contrast, the DIA envisions an updated regulatory system which is modern, flexible, and simple, as well as compatible with human rights, including freedom of expression. The review is currently in its third stage of identifying and refining options for a transformed content regulatory system.

The Content Regulatory Review

Next Steps for Review



⁵⁵ New Zealand Cabinet (2021), [Proactive release of Cabinet material about the initiation of the media content regulatory review](#)

⁵⁶ See: [Online Regulation Series 2.0: New Zealand](#).

For the purpose of the Review, ‘content’ is defined broadly as “any communicated material (for example video, audio, images, and text) that is publicly available, regardless of how it is communicated.” Meanwhile, harmful content is viewed on a spectrum, ranging from adult content to violent extremist content and child sexual exploitation material. The capacity of a piece of content to cause ‘harm’ depends on its audience, and the proposed classification is threefold:

- Content can cause harm to individuals by causing loss or damage to a person’s rights, property, or physical, social, emotional, and/ or mental wellbeing;
- Content can harm communities and identity groups, i.e. when members of a community experience harm because they are members of that community;
- Content can cause harm to wider society: This might look like individuals or communities losing trust in, or access to, public institutions such as the legal, health and education systems, their freedoms of identity and expression, and their right to democratic participation.

Key takeaways for tech companies:

At the time of writing,⁵⁷ the Content Regulatory Review was ongoing, and a proposed new framework was yet to be announced by the DIA. According to the timeline provided, a public consultation on the proposed new framework should have taken place between September and November 2022. Before any concrete announcement is made, it is difficult to assess the implications for tech platforms.

As part of the Review, the DIA has publicly released some provisional documents which are informing the deliberation of further proposals. These illuminate the essential principles engaged by the proposed framework.

- [Report: International Regulatory Frameworks for Online Content](#): the DIA commissioned a research study into current international frameworks regulating harmful content, comparing regulation case studies from Australia, Canada, Ireland, and the United Kingdom.
- This [Executive Digest](#) summarises the key issues, themes and concerns to consider in developing New Zealand’s own regulatory framework. The authors highlight some factors to consider in the specific context of New Zealand, in particular the explicit recognition of the risk of consolidating regulatory power in one government entity⁵⁸ and of the need to protect the safety and rights of minority, marginalised, and at-risk communities in New Zealand.⁵⁹
- [Cabinet material about the initiation of the media content regulatory review](#): The DIA has also released the relevant Cabinet papers and initiatory briefings circulated from February to July 2021.
 - The document hints at the scope of what is considered harmful content which includes ‘adult content’ such as pornography; violent extremist content, including material showing or promoting terrorism; child sexual abuse material; disclosure of personal information that threatens someone’s privacy; promotion of self-harm; mis/disinformation; unwanted digital communication; racism and other discriminatory content; and hate speech.⁶⁰

⁵⁷ November 2022.

⁵⁸ Thompson Peter and Michael Daubs (2021) [Executive Digest: International Regulatory Frameworks for Online Content](#), p.9

⁵⁹ Ibid, p. 16

⁶⁰ New Zealand Cabinet (2021), [Proactive release of Cabinet material about the initiation of the media content regulatory review](#), p.4

- o This is a wider scope than ‘objectionable’ content, which covers any publication that “describes, depicts, expresses, or otherwise deals with matters such as sex, horror, crime, cruelty, or violence in such a manner that the availability of the publication is likely to be injurious to the public good.”⁶¹ Given ‘objectionable’ publications are automatically banned under the Films, Videos, and Publications Classification Act (1993), it will be important for the eventual regulatory framework to differentiate between ‘harmful content’ and ‘objectionable content.’
- o The new regulatory framework will be “platform-neutral⁶² in its principles and objectives”, however it will also “need to enable different approaches to reaching these objectives, spanning Government, co-regulatory and self-regulatory approaches.”⁶³

The Aotearoa New Zealand Code of Practice for Online Safety and Harms

On 2 December 2021, [NetSafe](#), an online safety non-profit organisation, published the [Draft Aotearoa New Zealand Code of Practice for Online Safety and Harms](#), drafted in collaboration with leading tech platforms (including all founding members of the Global Internet Forum to Counter Terrorism (GIFCT), Twitch, and TikTok). NetSafe published the final [Aotearoa Code of Practice](#) on 25 July 2022. The voluntary Code “brings industry together under a set of principles and commitments to provide a best practice self-regulatory framework designed to enhance the community’s safety and reduce harmful content online.” The Code commits signatories to a set of outcomes and measures around seven themes of safety and harmful content: “1) child sexual exploitation and abuse 2) bullying or harassment; 3) hate speech; 4) incitement of violence; 5) violent or graphic content; 6) misinformation; and 7) disinformation.”

The Code creates an Administrator role and a multi-stakeholder Sub-committee, with the power to sanction by expulsion from the self-regulating community those signatories who fail to fulfil their commitments, as well as a public complaint mechanism and a regular review process every two years to consider amendments to the Code. The administrator role has been allocated to the [New Zealand Tech Alliance](#) (NZ Tech) – an advocacy organisation focused on promoting New Zealand’s digital technology industry.

Key takeaways for tech companies:

- Signatories to the Code are required to provide annual compliance reports to the Administrator, outlining progress in fulfilling their commitments, which will be made public and available for scrutiny.
- Signatories agree to make their “best efforts” towards four commitments outlined in the Code:
 - o Reduce the prevalence of harmful content online
 - o Empower users to have more control and make informed choices
 - o Enhance transparency of policies, processes and systems
 - o Support independent research and evaluation
- These broader commitments are divided into more specific outcomes (numbering 13 in total) and measures (45) that companies are expected to incorporate into their moderation practices. None of these relate to terrorism or violent extremism.
- The complaint mechanism permits New Zealand residents to complain if they believe signatory tech companies are not honouring their commitments.
- The Administrator and Oversight Committee have the joint power to remove a signatory for repeated breaches of the code.

⁶¹ New Zealand Department of Internal Affairs, [Objectionable and Restricted Material](#)

⁶² Covers all media channels (broadcasting, print, digital etc.)

⁶³ New Zealand Cabinet (2021), [Proactive release of Cabinet material about the initiation of the media content regulatory review](#), p.6



Tech Against Terrorism's Commentary

Familiar challenges for self-regulatory frameworks

Some civil society groups have warned that the Code is an attempt by larger tech companies to influence the legislative framework under consideration by the New Zealand government. Mandy Hank, CEO at Tohatoha NZ, a non-profit concerned with the social impact of technology, described the Code as “a weak attempt to pre-empt regulation - in New Zealand and overseas - by promoting an industry-led model.”⁶⁴ In practice, it is unlikely the code will significantly divert the development of regulation, especially given the explicit deference enshrined in the Code itself to any future legislation. The introduction of the Code can be understood within a broader international practice of adopting self-regulatory frameworks, including [Australia's Code of Practice on Disinformation and Misinformation](#) and the EU's [Code of Conduct on Countering Illegal Hate Speech Online](#).

Whilst the Code builds on existing self-regulatory frameworks, its authors [proclaim](#) it to be unique in its breadth and ability to hold tech platforms to account. However, in its current form the capacity of the Code and its provisions to deliver its stated impact are profoundly uncertain. As with similar legislative frameworks, “much of the code's operational substance is ambiguous or deferred to later processes”,⁶⁵ and its effectiveness depends heavily on the proper exercise of discretion by the Code's Administrator.

Critics argue that the administrator, NZ Tech, has insufficient capacity and expertise to effectively execute the complex role, which includes the assessment of signatories' transparency reports and determining whether there is a “proper basis” for rebuking non-compliance.⁶⁶ The Administrator's independence is also in question when the office is filled and funded by the signatories and is not superintended by formal bodies constituted under the code, such as a separately funded oversight board.⁶⁷ Tech companies considering signing up to the Code should consider how they can contribute to its effective implementation and improvement and apply its provisions to the development of measurable goals for tackling harmful content.

The Administrator and Oversight Committee do have the joint power to remove a signatory for repeated breaches of the code, and therefore do offer a means of enforcing tech platform accountability. The efficacy of this sanction lies in the embarrassment endured by tech companies in the public relations fallout consequent on their failure to meet commitments made as part of a voluntary self-regulatory mechanism. However, in practice this outcome is unlikely given the “unquantifiable and subjective assessments of compliance” and the vagueness of the threshold which must be met for a breach to be established, such as whether a signatory has given their “best efforts” to compliance.⁶⁸ Tech Against Terrorism recommends the Code's authors develop the commitments and outcomes further and in greater practical detail to provide more quantifiable objectives for signatories.

⁶⁴ Sawers Paul (2022) [Big Tech's push to self-regulate harmful content in New Zealand is 'weak attempt to preempt regulation', critics say](#), TechCrunch

⁶⁵ Curtis Barnes, Tom Barraclough, and Robins Allyn(2022) [Platforms Are Testing Self-Regulation in New Zealand. It Needs a Lot of Work.](#), Lawfare

⁶⁶ Ibid

⁶⁷ Ibid

⁶⁸ Ibid



SINGAPORE – UPDATE

Tech Against Terrorism first analysed Singapore’s online regulatory framework in the Online Regulation Series 1.0, shortly after the Protection of Online Falsehoods and Manipulation Act (POFMA) was passed in 2019. Since then, Singapore has continued to build its statutory basis for the regulation of online content, including the 2021 Foreign Interference Countermeasures Act (FICA). On 9 November 2022, the Singaporean Parliament passed its first legislation aimed at tackling harmful and illegal content online more generally, [the Online Safety \(Miscellaneous Amendments\) bill](#) (hereafter, OSB). This update focuses on the OSB and the requirements within it relating to the removal of terrorist content online removal, as well as to content moderation in general.

The Online Safety (Miscellaneous Amendments) Bill

In June 2022, the Minister for Communications and Information (MCI), Josephine Teo, announced an Online Safety Bill motivated by the “prevalence of harmful content” online. A public consultation, which closed in August 2022, followed the Minister’s announcement and the bill was passed by the Singaporean Parliament in November 2022 – a month after it was introduced in parliament. The bill came into effect on 1 February 2023, with the code of practice on online safety for designated social media services coming into force in the second half of the year.

The OSB introduces amendments to the existing [Broadcasting Act](#),⁷⁰ and of particular note is the requirement for platforms to comply with removal orders issued by the [Infocomm Media Development Authority](#) (IMDA), the country’s media regulatory authority. The amendments will target online communication services, defined broadly as services allowing communication or sharing of content for end users in Singapore, regardless of where the platform is based. Singapore’s OSB will also be supported by a Code of Practice for Online Safety, outlining requirements for designated social media services (SMSs) to limit the reach of harmful content. MCI has indicated that it will take an outcome-based approach to the Code, focusing on the results of the interventions taken by platforms rather than on specific actions.

Following the public consultation, [a summary](#) of the responses was published on the MCI website, showing that respondents were generally supportive of the bill whilst seeking safeguards for freedom of expression and privacy. Amongst the key points raised by the respondents together with industry groups, Tech Against Terrorism notes the following of interest to the discussion of regulation globally:

- Industry groups called for an outcome-based approach, considering the business model and size, as well as clarity on the thresholds for classifying designated services.
- Respondents agreed with the proposed safety features, although many admitted that they were not aware of existing measures developed by platforms and called for platforms to be more proactive in publicising them.
- Concerning IMDA orders, both respondents, including industry groups, highlighted the importance of ensuring there is an explanation as to why access to specific content should be disabled and deemed harmful.

⁷⁰ The Broadcasting Act, passed in 1994, is one of the key pillars of the media regulatory framework in Singapore, and is part of the foundational regulation empowering Singapore’s Infocomm and media regulatory agency, the IMDA.

Key requirements

- The IMDA will be empowered to adjudicate on “egregious content” and issue orders for social media platforms to disable access to such content.
 - “Egregious” harms include:
 - Content that advocates or instructs on self-harm, on violence including the physical torture of human beings, and on sexual violence Content depicting child nudity for a sexual purpose
 - Content that is likely to result in a public health risk
 - Content that promotes hostility or ridicule towards different racial or religious groups
 - Terrorist content, defined as “content that advocates or instructs on terrorism” – in relation to terrorist offences as defined in Singapore law”
 - Livestreamed videos of mass shootings and attacks⁷¹
 - The IMDA will be able to issue two types of blocking orders, requiring platforms to:
 - Block access to a specific content for Singaporean users
 - Block access to a specific account sharing egregious content to Singaporean users
- Platforms will be required to act on the IMDA removal orders “within hours”. The exact timeline for removal will depend on the content itself and the most egregious content, including terrorist content, are to be subject to shorter removal timelines.
- Platforms with significant reach will be designated as “Regulated Online Communication Services” (ROCS) and will be additionally required to comply with the Online Codes of Practices. The IMDA is yet to determine what constitutes a ROCS.
- The Online Safety Code will require designated SMSs to have “system-wide processes” to enhance online safety and have additional safeguards for users under 18s. The Code is expected to come into effect in 2023, following a consultation with the platforms concerned. [A draft code](#) and accompanying [guidelines on harmful content](#) have been published, requiring platforms to:
 - Have community standards and content moderation mechanisms in place to limit exposure to certain harmful content – Including: sexual violence, violent content, self-harm content, cyber-bullying, content endangering public health, and content facilitating vice and organised crime.
 - Offer tools for users to reduce their exposure to harmful content, for instance content filters for child accounts
 - Have easy-to-use reporting systems
 - Engage in proactive detection of CSAM and terrorist content
 - Publish an annual accountability report, detailing metrics on the effectiveness of their moderation processes.
- Under the bill, the IMDA can issue additional codes of practice and amend or revoke any codes of practice issued to ensure that the systems are in place and effective at addressing egregious content. Additional codes may also be put into place to provide practical guidance on content covered and procedures for collaborating with regulators or researchers. However, before any new code is introduced, IMDA is legally bound to consult ROCSs who may be subject to the code and obtain their feedback on the code
- Failure to comply with a code can lead to the service providers being fined up to approx. \$725k (SGD1 million) or being blocked entirely in Singapore by order of the IMDA.

⁷¹ Livestreamed terrorist acts were amongst the examples given by the MCI when introducing the law in June, in particular relating to how harmful content can be amplified on social media.



Tech Against Terrorism's commentary

A co-regulatory approach

The OSB and its Code of Practice highlight Singapore's co-regulatory approach to the online sphere. Singapore is home to the Asia-Pacific headquarters for most of the leading global tech companies, and has collaborated closely with tech platforms in building its regulatory framework in order to "develop regulations fit for Singapore's socio-cultural context". This explicit co-regulatory approach emphasises that regulations can be developed with rather than against platforms and has the potential to promote a local approach to global content moderation standards set by the tech sector. However, Tech Against Terrorism calls on policymakers to be careful not to consider solely the perspective of large tech companies. Policymakers should ensure that small, medium, and newer platforms are given an equal voice as their capacity to comply with online regulation will greatly differ from those of platforms with greater resources.

Exclusion of private messaging

Contrary to the POFMA which applies equally to public facing social media and to private messaging services, the OSB excludes private messaging from its scope. This was discussed in Parliament when the law was passed, with Minister Teo explaining that "legitimate privacy concerns" led to this exclusion.⁷² Minister Teo highlighted that there are other resources available to users to block or report unwanted messages to the service providers. However, Minister Teo also specified that groups with large membership will be considered public-facing platforms and will be in scope. At the time of writing,⁷³ it is not clear whether online services offering both private communications and large groups will be covered, nor how such "large groups" will be defined.

Tech Against Terrorism commends this exclusion of private communication platforms, which safeguards the fundamental right to privacy. User reporting is an underestimated yet critical tool to counter the diffusion of illegal and harmful content online when shared on private messaging services, including those offering end-to-end encryption (E2EE). For a review of user reporting features available on encrypted messaging services, please see Tech Against Terrorism's report on ["Terrorist Use of E2EE: State of Play, Misconceptions, and Mitigation Strategies"](#).

Undefined scope

At the time of the law's adoption, November 2022, there were a number of uncertainties regarding its final scope and form. It is unclear how the IMDA will define services of "significant reach", and therefore which platforms will have to comply with the Code of Practice.

The concept of "egregious" content is also difficult to define precisely and therefore to operationalise, because such content exists in a 'grey area' – a point considered in the course of Parliamentary discussions of the bill. Whilst the bill clearly lists what type of content can be considered egregious, not all included forms of harm refer explicitly to harms accounted for in Singapore's existing legal framework – contrary to the listing of terrorist content which refers to Singapore's counterterrorism framework. Tech Against Terrorism calls on policymakers to ensure that prohibition of online content is based on existing legal framework, and to refrain from making what is legal offline illegal online.

⁷² Tham Irene (2022), [New rules to make social media firms accountable for online harms](#), The Straits Times

⁷³ November 2022



UNITED ARAB EMIRATES

Both internet access and content are heavily restricted in the UAE, resulting in Freedom House's 2022 designation of the UAE as "not free".⁷⁴ The UAE has a multi-layered regulatory framework that criminalises the use of the internet for, amongst other things, terrorist purposes. The definition of terrorism in UAE legislation is both broad and vague, encompassing anything from content produced by or promoting terrorist organisations, to content that threatens the stability, safety, security, or unity of the UAE.

The UAE's regulatory framework

- **Federal Law No. (2) of 2006 on The Prevention of Information Technology Crimes:** The Law criminalises the use of the internet for a range of purposes, including terrorism, identity fraud, money laundering, and pornography. It covers digital media such as blogs, SMS, emails, and social media.
 - Article 21 imposes a penalty of up to five years imprisonment for whoever sets up a website or publishes online information on behalf of a terrorist group, uses pseudonyms to facilitate contact with its leaders or members, promotes its ideas, provides financing to the group or publishes information related to the manufacturing of incendiary, explosive and other devices used in terrorist activities.
- **Federal Law Number 5 of 2012 on Combatting Cybercrimes:** The Law amended the previous 2006 Law. It provides a legal basis to prosecute and jail people who use information technology for the wide range of purposes stated above.
- **Federal Law No. (7) of 2014:** The Law sets out every act that the UAE deems a terrorist offence (capable of leading to a 'terrorist result') and the relevant sanctions for those offences.
 - The Law defines 'terrorist result' as an action committed for the purpose of "inciting fear among a group of people, killing them, or causing them serious physical injury, or inflicting substantial damage to property or the environment, or disrupting the security of the international community, or opposing the country, or influencing the public authorities of the country or another country or international organisation while discharging its duties, or receiving a privilege from the country or another country or an international organisation."
- **Federal Decree Law No. 2 of 2015 on Combatting Discrimination and Hatred:** The Law criminalises discrimination or hate speech against individuals or groups based on race, ethnic origin, religion, doctrine, and so on.
 - The Law applies to the internet, telecommunication networks, electronic websites, amongst others, criminalising discrimination and/or hate speech online.
- **The Sawab Centre, March 2015:** Headquartered in Abu Dhabi, the Sawab Centre is a partnership between the UAE and US that aims to counter extremist propaganda and terrorist messaging in the online space.⁷⁵

⁷⁴ Freedom House (2022), [Freedom in the World 2022: The United Arab Emirates](#).

⁷⁵ Embassy of the United Arab Emirates Washington DC, [Counterterrorism](#).

- **Internet Access Management (IAM) Regulatory Policy, 2017:** The regulation aims to “ensure the security of the internet and protect end-users from harmful websites” by prohibiting categories of online content. The IAM is implemented by the Telecommunications and Digital Government Regulatory Authority.
 - The policy prohibits terrorism, which includes internet content that relates to terrorist groups, facilitates communication with leaders or members, attracts members, promotes or favours their ideas, assists in financing their activities or making devices to be used in attacks.
 - The IAM also prohibits under its more general prohibition of ‘terrorism’ “Internet content that incites, encourages or enables the commission of an act that would intend to threaten the stability of the UAE or its safety, unity or security, or oppose the basic principles underlying the regime, or intend to overthrow or takeover the regime”.
- **Federal Decree Law No. 34 of 2021 on Combatting Rumours and Cybercrimes:** A reform of the repealed Federal Law by Decree No. 5 of 2012 on Cybercrimes. It imposes heavy fines for storing and sharing illegal content, which includes terrorist content.
 - Article 53 of the Law imposes a fine ranging between 300,000 and 10,000,000 Dirhams (equating to approximately 66,000 to 2,200,000 GBP) on any individual that uses the internet or an electronic account to store or share ‘unlawful content’ or refrains from removing said content.

Main regulatory bodies

- **The Telecommunications and Digital Government Regulatory Authority (TDRA):** Federal Law No. 3 of 2003 Regarding the Organisation of the Telecommunications Sector established the TDRA as the regulator of the telecommunications and information technology sector in the UAE.
 - The TDRA monitors online content available to users in the UAE and notifies website operators based in the UAE of any users potentially in breach of the IAM policy.
 - The TDRA uses the state-owned internet service providers (ISPs) Etisalat and Du (see below) to enforce its regulations and uses anti-terrorism laws to prosecute individuals who breach those regulations.⁷⁶

Other relevant organisations

- **The Emirates Telecommunications Corporation or Etisalat:** The largest telecom provider in the UAE. Etisalat acts under instruction from the TDRA to enforce online regulations.
 - From 1976 to 2006, Etisalat was the sole telephone and telecoms provider in the UAE.
- **Du:** The UAE’s second telecom operator, rival to Etisalat. Du has a monopoly as the telecom provider in UAE ‘free zones’.
 - In 2008, Du announced that it would begin blocking sites that conflict with the UAE’s “moral, social and cultural values”, in line with TDRA regulations.⁷⁷

⁷⁶ Porutiu Theodor (2022), [Censorship in the UAE: How to Get Around it](#), VPN Overview.

⁷⁷ Noueihed Lin (2008), [UAE’s du starts restricting Internet access](#), Reuters.



Key Takeaways for Tech Platforms

The UAE enforces its regulations through the state-owned internet service providers Etisalat and Du. Tech platforms are not responsible for removing prohibited content, but access to their services, or content on their services, may be blocked for users in the UAE if it is found to be in violation of UAE law. Fines and criminal punishments are given to individual users, rather than tech platforms.

TAT Analysis and Commentary

State control and restrictions of freedom of expression

Human and digital rights advocates have raised strong concerns around laws in the UAE, with Article 19 stating that UAE law “severely threatens and unduly restricts the right to freedom of expression both online and offline.”⁷⁸

In particular, Article 19 criticised the “overbroad and vague terminology used” in the 2021 Cybercrime Law, with specific reference to the invocation of ‘national security’ which provides the authorities “with excessive discretion to criminalise and impose lengthy prison sentences on individuals exercising their right to freedom of expression”.⁷⁹ According to Human Rights Watch, the incorporation of vague terminology into the law enables the UAE to prohibit any kind of online speech that may be critical of the UAE, its authorities, and its leaders, effectively banning “any speech made in online public forums or in private chats that the government does not approve”.⁸⁰ Concerns over the ambiguous and unclear nature of the law have been echoed by the UN Special Rapporteur on the Independence of Judges and Lawyers, Gabriela Knaul, who has highlighted its violation of international human rights standards.⁸¹

The UAE’s broad definition of “[t]errorist result” under its 2014 Terrorism Law likewise enables its government to ban and punitively sanction a wide scope of online speech. For instance, in February 2015 UAE authorities arrested the al-Suwaidi sisters for tweeting about their brother, Mr. Eissa al-Suwaidi, who had been sentenced as part of the unjust UAE94 trial.⁸² The Emirati authorities later responded to the International Centre for Justice and Human Rights’ investigation with the assertion that the sisters had been imprisoned because their brother belonged to a “terrorist organisation”, according to article 22, paragraph 2, of the 2014 terrorism law.⁸³ Human Rights Watch has further criticised the UAE for allowing “any act that courts deem to have antagonised the state, stirred panic, or undermined national unity to be designated as terrorism.”⁸⁴

By mandating Etisalat and Du to block access to non-compliant platforms, the UAE’s laws establish platform blocking as an integral part of the country’s content regulation strategy. According to official statistics, 883 websites were blocked in the first quarter of 2022.⁸⁵

⁷⁸ Article 19 (2022), [United Arab Emirates: New cybercrime and anti-rumour law violates rights](#).

⁷⁹ Article 19 (2022), [United Arab Emirates: New cybercrime and anti-rumour law violates rights](#)

⁸⁰ Human Rights Watch (2022), [UAE: Sweeping Legal ‘Reforms’ Deepen Repression](#)

⁸¹ International Centre for Justice and Human Rights (2016), [ICJHR submission to the Office of the UN High Commissioner for Human Rights on the negative effects of terrorism on the enjoyment of all human rights and fundamental freedoms in the United Arab Emirates](#)

⁸² Amnesty International (2021), [UAE: Nearly a decade of unjust imprisonment for ‘UAE-94’ dissidents](#)

⁸³ International Centre for Justice and Human Rights (2016), [ICJHR submission to the Office of the UN High Commissioner for Human Rights on the negative effects of terrorism on the enjoyment of all human rights and fundamental freedoms in the United Arab Emirates](#)

⁸⁴ Human Rights Watch (2014), [UAE: Terrorism Law Threatens Lives, Liberty](#)

⁸⁵ Khaleej Times (2022), [UAE: 17 categories of online content, websites that are blocked in the emirates](#)

Moreover, both Etisalat and Du, the only entities responsible for providing internet access to the UAE, are either directly or indirectly owned by the state. As a result, the UAE government has control in ensuring all internet usage in the UAE abides by its law and the regulations of the TDRA. The TDRA can order Etisalat and Du to remove online content without judicial oversight, often issuing takedown notices or enacting website blocking before any criminal complaint has been made to the police.⁸⁶

The practice of website blocking contravenes international human rights law, according to AccessNow.⁸⁷ In particular, the UAE's blanket bans on entire communications services - such as voice-over-internet-protocol (VoIP) service Skype, and the voice calling function of WhatsApp – do not satisfy UN Human Rights Special Rapporteur David Kaye's criteria for the promotion and protection of the rights to freedom of opinion and expression.⁸⁸

⁸⁶ Porutiu Theodor (2022), [Censorship in the UAE: How to Get Around it](#), VPN Overview.

⁸⁷ Access Now (2017), [Access Now submission to the Universal Periodic Review: UAE, Third Cycle](#).

⁸⁸ United Nations Office of the High Commissioner for Human Rights (2017), [Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression](#).



UNITED KINGDOM – UPDATE

Tech Against Terrorism analysed the United Kingdom’s regulatory framework in the Online Regulation Series 1.0 in 2020. Following that first analysis, the UK Department for Digital, Culture, Media & Sport (DCMS) published the draft Online Safety Bill (OSB) in May 2021, which aims to counter illegal content online. This 2022 update focuses on the status and implications of the Online Safety Bill as it pertains to the regulation and moderation of online content, including terrorist content.

Editorial note: This entry was last edited in December 2022, the legislative process for the Online Safety Bill has made further progress since then and was moved for discussion by the House of Lords in January 2023.

The Online Safety Bill

Though the final form of the Online Safety Bill and its practical implication for tech platforms remain uncertain, the Bill in its current form emphasises protecting children in the internet ecosystem, preventing the spread of terrorist propaganda, and reducing the prevalence of illegal content. The OSB will appoint the United Kingdom’s communications regulator, Ofcom, as the enforcement agency for this new ‘duty of care’ imposed on entities falling within scope of the legislation.

Timeline:

- April 2019: UK Government issues its Online Harms White Paper, followed by a three-month public consultation period
- May 2021: [Draft Online Safety Bill](#) introduced
- July 2021: House of Lords committee publishes [report on freedom of expression online](#)
- March 2022: Bill introduced to Parliament by Culture Secretary Nadine Dorries
- May 2022: [First debate in House of Commons](#)
- June 2022: [Committee amended Bill](#) published
- July 2022: [Debate of amended Bill in Commons](#)
- July 2022: Online Safety Bill delayed due to leadership changes
- October 2022: [Lords debate on their July 2021 report](#)

The Online Safety Bill applies to any digital user-to-user service which enables users to generate, upload, and share content, and to encounter content shared by other users, and also applies to search services allowing users to search more than one website or database. This includes public and private channels, search engines, content-sharing platforms, social media platforms, blogs, forums, listings sites, aggregators, and any provider that allows one user to encounter content from another user. As it currently stands, services governed by the bill will be obliged to deal with illegal content. Ofcom will have additional powers to ensure companies take actions to tackle terrorist activity and child sexual abuse and exploitation online, including the power to require tech companies to disclose how they deal with harms and to take enforcement action against companies failing to fulfil their obligations. In November 2022, the DCMS and Secretary of State for Digital Issues Michelle Donelan announced several proposed changes,⁸⁹ including new duties to protect freedom of expression online, in response to widespread concerns that the bill would lead to the over-removal of content.

⁸⁹ Government of the United Kingdom – Department for Digital Culture, Media & Sport (2022), [New protections for children and free speech added to internet laws](#)

Regarding terrorist content, the OSB in its current form⁹⁰ would require platforms to:

- Use proportionate measures to mitigate and manage the risks of harm to individuals, and the impact of harm, as identified in illegal content risk assessments. Platforms should also have systems and processes in place to prevent users from encountering illegal or harmful content.
- Use proportionate systems and processes to remove illegal content when notified or otherwise made aware of it.
- Specify clearly in terms and conditions how individuals are protected from illegal content, addressing terrorism specifically, how terms and conditions are applied consistently, and how technology is used proactively to identify illegal content.
- Use systems and processes that allow users and other affected persons to report illegal and harmful content.
- Establish accessible procedures by which users can appeal decisions to remove or restrict content.

The requirements in the Bill should be proportionate and reflect the size, resources, and risk level of companies in scope though it remains to be seen how these variables are defined. Large and high-risk platforms, for instance, will have to clearly stipulate in their content standards what types of legal content adults can post. However, the threshold of what constitutes “large and high-risk” platforms is yet to be determined by DCMS in consultation with Ofcom.

Tech Against Terrorism’s Commentary

Freedom of expression and the right to privacy

Tech Against Terrorism commends proposed amendments to the bill which seek to protect the freedom of expression online, and which mitigate the risk of making what is legal offline illegal online. While the bill still considers false and threatening communications an offence, the revision includes the removal of “harmful communications” as an offence, which risked a chilling effect on the freedom of expression. The proposed amendments also include a requirement for major platforms to not remove content that does not violate the law or does not clearly breach the Terms of Service.

However, the broad scope of the Bill and the lack of precise or operable definitions of terrorist content generally in the current draft means that platforms may have to decide for themselves what content is ‘justiciable’ and by what standard the lawfulness of that content should be determined. These standards and definitions will vary hugely across the tech sector ecosystem, depending on the resources available for moderation, and may come to appear arbitrary. Deciding what is legal or illegal online is the duty of democratic institutions and independent judicial bodies, not private tech companies. As it stands, the current definition of “democratic” and “journalistic” content will be difficult to implement. Without operable definitions and clear thresholds for the removal or protection of content, platforms risk over-removing content and inhibiting users’ freedom of expression. There is also a risk of under-removal of terrorist and violent extremist content disguised as “democratic” and “journalistic”. Tech Against Terrorism calls for caution around terrorist and violent extremist actors exploiting this particular dispensation. To adequately identify terrorist and violent extremist content, without infringing on users’ rights, platforms need clear and operable definitions of what is legitimately journalistic and political content and what is not and why, as well as practical support to correctly identify TVE content feigning legitimacy.

⁹⁰ As of December 2022



The UK OSB is set to apply indiscriminately to private communication and public content online, thus infringing on the fundamental right to privacy online. Interferences with the right to privacy should be proportionate to the stated aims.⁹¹ This is not the case with a monitoring mandate which would compromise the security and privacy of all UK online users with limited effect on criminal actors who are likely to migrate to non-compliant platforms. Tech Against Terrorism published an in-depth report assessing the risk of terrorist use of end-to-end encrypted services, outlining possible mitigation strategies that safeguard encryption and the right to privacy. You can find the report [here](#).

Mandating risks assessments

Commendably, the bill encourages tech platforms to undertake risk assessments. Understanding the risk is a crucial first step to effectively countering terrorist exploitation of the internet, and tech companies should to the best of their abilities consider how their platforms could be exploited and remain aware of adverse usage. However, smaller platforms will need support in carrying out risk assessments, as they may not have the resources or capacity to conduct these. In our view, additional resources should be made available to enforce existing legislation capable of tackling illegal and harmful content, and to prosecute those creating and sharing that content before enacting a new law which risks fragmenting the regulatory landscape.

⁹¹ United Nations General Assembly [Resolution 68/167](#)



UNITED STATES – UPDATE

Tech Against Terrorism first analysed the United States’ regulatory framework in the first edition of the Online Regulation Series in 2020. Since the first analysis, six pieces of state and federal legislation have been passed across the US with considerable implications for content moderation and counterterrorism efforts online. Calls for the regulation of online user-generated content, including by means of reforms to Section 230 of the Communications Decency Act, continue to be made by Democrats and Republicans at the state and federal level.

This 2022 update focuses on several pieces of legislation both enacted and proposed across the United States as it pertains to the regulation and moderation of online content. For a more complete overview of the US regulatory landscape, please also refer to the dedicated entry in the first edition of the Online Regulation Series Handbook.⁹²

Regulation mandating increased transparency

In September 2022, California Governor Gavin Newsom signed AB 587 into law, establishing new transparency requirements for social media companies. The new requirements include the publication and submission, to the California Attorney General, of bi-annual reports on company content moderation practices. The law applies to “social media companies”, defined as persons or entities that own or operate one or more social media platforms with users in California, and with gross revenues of over 100 million USD.

The bi-annual reports, which must be published from 1 January 2024, must include:

- Information on how the terms of service define certain categories of content (e.g., hate speech, extremism, disinformation, harassment and foreign political interference)
- How automated content moderation is implemented
- How the company responds to reports of violations of the terms of service
- How the company responds to content or persons violating the terms of service

The reports must also provide detailed metrical analysis of flagged content, including:

- The number of flagged items
- The categories of flagged content
- The number of times flagged content was shared and viewed
- Whether action was taken by the social media company, and if so, what action

Social media companies that do not supply their terms of service to the California Attorney General’s office by 1 January 2024 or fail to submit their reports or materially omit required information from their reports are subject to fines of up to 15,000 USD per violation per day. Thereafter, reports will be due biannually on April 1 (covering the third and fourth quarters of the previous year) and October 1 (covering the first and second quarters of the year).

For social media companies, compliance with the reporting requirements will require collecting and managing information that may not have previously been subject to such attention. Tech Against Terrorism has published Transparency Reporting Guidelines outlining frameworks for transparency which can be adapted by platforms of different sizes and service offerings.⁹³ The guidelines outline a number of core metrics to report on that cover policies, processes, systems, and outcomes.

⁹² Tech Against Terrorism (2021), [The Online Regulation Series Handbook](#), p.128

⁹³ Tech Against Terrorism (2021), [Guidelines on Transparency Reporting on Online Counterterrorism Efforts](#)



Regulation addressing platforms' ability to moderate content

Texas House Bill 20

On September 16, the Fifth Circuit of the US Court of Appeals issued its decision in *NetChoice L.L.C. v. Paxton* upholding Texas House Bill 20, which is a law that limits the ability of large social media platforms – meaning platforms with more than 50 million active users in the US and with users in Texas – to moderate content. NetChoice is likely to ask the U.S. Supreme Court to review the Fifth Circuit's decision.

Entities covered by the Texas House Bill 20 are obliged to:

- Prohibit “censor[ing] a user, a user’s expression, or a user’s ability to receive the expression of another person” based on the “viewpoint” of the user or another person, or the user’s location
- Publish information about their algorithms for displaying content
- Publish an acceptable use policy with information about their content standards and restrictions
- Provide users with an explanation for each decision to remove their content, as well as a right to appeal the decision

The Court concluded that because online services exercise limited editorial control, most of them doing so via algorithmic recommendations, they cannot enjoy the same protections afforded under the First Amendment to newspapers, for example.

The court further explained that platforms have the ability to create original content on their service to “say whatever they want to distance themselves from the speech they host.” However, the Court has not considered Section 230, which provides legal immunity to tech companies in respect of their decisions to remove content shared by third-party “publishers or speakers” on their website. As such, this case will likely be reviewed by the US Supreme Court.

Ohio House Bill 441

The state of Ohio has passed legislation prohibiting social media companies of over 50 million users from “censoring”⁹⁴ their users.

The effect of the legislation is to:

- Prevent companies from removing posts or expelling people from their platforms based on the “viewpoint” of users or the ideas expressed by users in their posts. The law does not protect speech already illegal under federal law like harassment or incitement to violence.
- Allow private citizens to sue social media companies if their views are “censored.”
- Declare that social media companies are “common carriers” — which do not receive speech protections as publishers in the way that newspapers do. This point is currently being disputed as social media companies regularly screen, moderate, emphasise and curate content, which are activities consistent with publishing.

⁹⁴ This includes blocking, banning, demonetising, deplatforming, removing, denying “equal access or visibility to,” or otherwise “discriminating” against the user based on what they post.



Florida's Senate Bill 7072

In Florida, Senate Bill 7072, which prohibits platforms from banning or deprioritising candidates for state office, and places similar prohibitions on news outlets above a certain size, has passed in the Senate.

Unlike in Texas, Florida's bill seeks to address the concept of "deplatforming" individuals, whereas Texas' bill addresses social media companies moderating users' content on their platforms. A US District Court which examined the Florida law found that social media companies were protected by the First Amendment when making content moderation decisions. Giving their ruling, the judges concluded that social media platforms' content moderation activities constitute 'speech' within the meaning of the First Amendment. Consequently, Florida asked the Supreme Court to review the court ruling.

The law would expose social media companies to lawsuits when users or state bodies determine that content or user accounts have been moderated in a manner violating

Update on Section 230

Section 230 of the Communications Decency Act of 1996, a federal statute, continues to be interpreted expansively in practice. The clause, which establishes exemptions from liability for intermediaries in connection with user-generated content, essentially means that tech companies are not liable for content posted by their users.

The US Supreme Court approved in October 2022 a request to review the validity of Section 230, subjecting the scope of Section 230 to review by the Court for the first time. The case, *Gonzalez v Google*, arises from the 2015 terrorist attacks in Paris, in which the plaintiff Reynaldo Gonzalez's daughter, Nohemi Gonzalez, was murdered by ISIS terrorists. Her family members brought an action against Google, alleging that the company had aided and abetted ISIS by hosting their videos. Although simply hosting a video would normally be protected by Section 230, the plaintiffs additionally alleged that YouTube had affirmatively recommended ISIS videos, which helped ISIS recruit new terrorists. The District Court disagreed and dismissed the claim, the Ninth Circuit dismissed the Gonzalez' appeal, and the case is currently awaiting a substantive hearing at the Supreme Court.

Both the Trump and Biden administrations have expressed a desire to amend Section 230. In May 2020, President Trump released an executive order asking regulators to redefine Section 230 more narrowly and asked agencies to amass existing complaints of political bias that would justify revoking the legal protections offered by Section 230. In January 2020, Biden proposed revoking Section 230 entirely.



Tech Against Terrorism's commentary

Online regulation in the US is defined by a recent pattern of content moderation laws emerging at the state level. Oklahoma Senate Bill 383 stipulates that users can sue social media platforms if they censor political or religious speech. Idaho House Bill 323, also passed this year, similarly gives users a cause of action against social media companies that “censor” them. West Virginia House Bill 3307, passed only by the lower house so far, would prohibit censorship of West Virginia political candidates on social media platforms by the companies that control those platforms. Similar legislative proposals, such as Senate Bill 111 in Kentucky and Legislative Bill 621 in Nevada, as well as the reintroduction of a bill in North Carolina which would limit Section 230 immunity to ‘Good Samaritans’, all emphasise that the increasing attempts to rein in social media platforms are motivated by the disputed concern that such companies are too large and too liberal.

This trend of hostility to tech companies invites reconsideration of the responsibility digital services have to protect the communities whose content they host from online harm, including exposure to racism, abuse, and subversive propaganda tolerated in the name of permitting “the free exchange of ideas”. To oblige private companies not to differentiate between harmful and benign online content, by treating statements such as “God Bless America” and “Death to America” equally, is unconstitutional. Such an obligation impacts considerably on the prevalence of terrorist and violent extremist rhetoric online. Terrorists and violent extremists often exploit legal clauses in countries around the world that permit the sharing of terrorist content for journalistic or research purposes. Violent far-right extremists, for example, often share graphic content or instructional material alongside a deliberate caveat that they are sharing for ‘journalistic’ purposes, and that they ‘do not endorse’ the material being shared. Such deliberate misrepresentation is a tactic intended to circumvent both automated and human content moderation. Furthermore, many TVE actors intentionally dilute their rhetoric to avoid de-platforming despite their rhetoric being overtly supportive of violence elsewhere. This is a particular tactic of violent far-right extremists, who pose as political commentators on mainstream platforms and exploit that association and implied approval to secure legitimacy while posting more extreme content elsewhere.



GENERAL UPDATE

In this entry, we provide an overview of recent regulatory development in jurisdictions previously covered in the Online Regulation Series 1.0 and 2.0. In doing so, we aim to provide an update on ongoing regulatory discussions when bills have not yet been passed into law.



Canada

In January 2021, the then Heritage Minister Steven Guilbeault announced that the Canadian government would introduce a new regulatory framework for online service providers. This announcement was followed by the publication of [a technical paper](#) outlining the key tenets of Canada's future bill on addressing harmful content online, accompanied by the release of explanatory notes on the bill and the launch of a public consultation. As of November 2022, the Bill was being discussed and revised by the Canadian government, you can find more about the key principles of regulations outlined in the technical paper in the [Online Regulation Series 2.0](#).⁹⁵ The below update focuses on the salient issues identified in the public consultation conducted by the Canadian government.

2021 Public Consultation

The Canadian Government published a summary of the 2021 public consultation, titled "What we've heard", highlighting a number of key concerns with the proposed framework. Those concerns include:

- The wide range of regulated entities
- The types of harmful content covered
- The proactive monitoring obligations and 24h take-down period for Online Communication Service Providers (OCSPs)
- The practical provisions for the blocking and filtering of online content
- The mandatory reports of user information to law enforcement for certain classes of content such as terrorist material

2022 Expert group

This feedback led the Canadian government to launch a new phase of consultation under the form of an Expert Advisory Group. This group launched in March 2022 and concluded in June of the same year and was tasked with examining the different components of the proposed legislation and how to incorporate into the proposal amendments suggested during the public consultation. The Expert Advisory Group comprised 12 individuals who participated in 10 weekly workshops addressing 8 different elements of the regulatory framework:

1. Subjects of regulation
2. Objects of regulation
3. Legislative and regulatory obligations
4. Regulatory powers
5. Risk-based approach
6. Freedom of expression and other rights
7. Connection to law enforcement
8. Disinformation

⁹⁵ Tech Against Terrorism (2021), [The Online Regulation Series – Canada Update](#).

A summary of the workshop discussion was published after each session. The [summary for the concluding workshop](#) contains the highlights from the different discussions that took place, including areas of agreement and disagreement amongst the experts. The disagreements are particularly interesting for emphasising the difficulties of building a regulatory framework that accounts for the changing online threat landscape and the challenges inherent both in establishing a definition of what constitutes harmful content and in identifying efficient mitigation strategies that do not infringe on fundamental rights. Some of these points are reminiscent of those raised by Tech Against Terrorism in the previous editions of the Online Regulation Series,⁹⁶ including platforms' liability, inclusion of private communication services within the scope of regulatory oversight, consideration for the size and resources of platforms, definition of harms, considerations for freedom of expression and the risks of over-removal of content.



Brazil

In March 2022, the government of former president Jair Bolsonaro presented a revised version of the Brazilian Internet Freedom, Responsibility and Transparency Act, also referred to as the '[fake news bill](#)'. The original proposed bill received significant backlash by civil society groups for provisions which infringed on users' privacy and security.⁹⁷ The bill was approved by the Federal Senate in June 2020 however it has since gone through several revisions. The most recent iteration of the bill has dropped its focus on criminalising online speech and traceability, which broadly posed a risk to platforms, including E2EE services. Instead, the bill focuses on platform transparency, specifically requesting platforms to be public about the size of their content moderation teams and the publication of metrics around false content removal. In April 2022, the Chamber of Deputies voted against fast-tracking the bill and as result must go through a longer legal process before it can be passed.⁹⁸

As highlighted in our previous update on Brazil's online regulation, there has been significant public debate around the proposed presidential decree on content moderation, which used the guise of freedom of expression to restrict platforms' abilities moderate content.⁹⁹ Since the Federal Senate opposed President Bolsonaro's initial decree, the government attempted to re-purpose it as a 'Provisional Measure', which still allowed the government to give out financial penalties to tech platforms for content moderation. Whilst this was also met with backlash and eventually rejected, in September 2022, the government tried for a third time to push the same legal provisions but under a different bill.¹⁰⁰ As of November 2022, there had been no significant progress made on the passing of this bill.

⁹⁶ It should be noted similarities in topics here does not equal agreement.

⁹⁷ For an analysis of the original draft law on Brazilian Internet Freedom, Responsibility and Transparency Act, see the first edition of the [Online Regulation Handbook](#), p.142.

⁹⁸ Freedom House (2022), [Freedom on the Net 2022 – Brazil](#)

⁹⁹ For more information on the Presidential decree on content moderation see the original proposal, <https://static.poder360.com.br/2021/06/minuta-decreto-mudanca-marco-civil-internet.pdf>

¹⁰⁰ Internet Society (2022), [Internet Impact Brief: Proposals to Regulate Content Moderation on Social Media Platforms in Brazil](#)

Indonesia

Tech Against Terrorism analysed Indonesia's regulatory framework in the Online Regulation Series 2.0 in November 2021.¹⁰¹ Our first analysis focused on the Ministerial Regulation 5 (MR5), which came into effect in 2020, as well as the Law No.11 on Electronic Information and Transaction (2008) and its amendment, and the Law No. 19 (2016).

Since November 2021, the Indonesian government has made no significant amendments to the MR5, despite being subject to heavy criticism on account of its breach of international human rights obligations.¹⁰² The MR10, the only amendment to MR5, was passed on 21 May 2021 and that amendment consisted solely of the insertion of article 47, imposing an obligation on private electronic system operators (ESO) to register within six months of Indonesia's Online Single Submission system becoming operational.¹⁰³

In March 2022, Reuters reported that Indonesia was preparing stricter measures under the MR5.¹⁰⁴ The new measures would include fining platforms per item of objectionable content, with fines rising the longer content stays on platforms and potentially reaching millions of rupiah (1m rupiah equating to approximately 65USD). The strict measures also establish criminal liability for employees of platforms who fail to comply with government requests on too many occasions, the number of which is unspecified. Such measures have not yet entered into effect as of the end of this coverage period.

On 22 June 2022, Kominfo¹⁰⁵ announced a short deadline of 20 July 2022 for tech platforms to register themselves pursuant to the MR5 and MR10. This included social media companies, search engines, messaging services, mobile applications, and in effect most other online services and applications. Those who failed to comply risked having their services blocked in Indonesia.¹⁰⁶ 10 days after the deadline, Kominfo ordered the blockage of eight electronic service operators (ESOs) for failure to comply - including Electronic Arts' Origin, Epic Games, PayPal, Valve Software (Steam and published games Counter-Strike: Global Offensive and Dota 2), Xandr, and Yahoo. The platforms were later unblocked after their registration under the MR5.

Despite considerable backlash on social media, with many individuals and civil society groups sharing the hashtag 'BlockirKominfo' on Twitter, Samuel Pangerapan, Director General of ICT Applications at the Ministry of Communication and Information Technology, dismissed criticism with the assertion that the measure would help protect Indonesia's internet users.¹⁰⁷

¹⁰¹ Tech Against Terrorism's Online Regulation Series 2.0, [Indonesia](#)

¹⁰² Access Now (2022), [Global coalition of NGOs urge Indonesia to repeal censorship regulations](#)

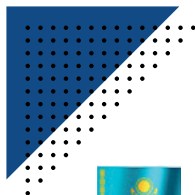
¹⁰³ Indonesia, Ministry of Communication and Informatics (2021), [Regulation of the Minister of Communication and Informatics Number 10 of 2021 concerning Amendments to the Regulation of the Minister of Communication and Informatics Number 5 of 2020 Concerning Private Electronic System Operators](#)

¹⁰⁴ Potkin Fanny and Sulaiman Stefano (2022), [Indonesia preparing tough new curbs for online platforms](#), Reuters.

¹⁰⁵ Kominfo, The Ministry of Communication and Information, are responsible for information and communication affairs in Indonesia.

¹⁰⁶ Caster Michael (2022), [Internet Freedom in Indonesia is Teetering on a Razor's Edge](#), The Diplomat.

¹⁰⁷ Bonifacig Igor (2022), [Indonesia blocks Steam, PayPal and other services over missed regulatory deadline](#), Engadget.



Kazakhstan

In May 2022, the government of Kazakhstan enacted amendments to the law on the protection of the rights of the child which began as a proposal to ensure foreign social media companies had a physical presence within the country. The government of Kazakhstan saw this as an opportunity to exact more control over foreign companies operating platforms internationally, specifically for the purposes of content removal. The original proposal in September 2021 was highlighted in our previous analysis of Kazakhstan’s online regulation.¹⁰⁸

Before coming into force in May 2022, however, it went through several iterations and some of the harsher provisions were removed, such as the total blocking of services which failed to establish a base in Kazakhstan or failed to remove content that was deemed illegal by the government. As reported by Freedom House, the Parliament of Kazakhstan’s lower chamber restricted the power initially given to the Ministry of Information and Social Development (MISD) to directly order the blocking of social media platforms, messaging applications and websites without obtaining a court order. Nevertheless, under the new law, MISD are permitted to directly request the removal of illegal content, which is defined broadly in Kazakh law.¹⁰⁹



Poland

As highlighted in the Online Regulation Series 2.0,¹¹⁰ the Polish government was expected to pass the ‘Act on the Protection of Freedom of Speech on Social Networking Sites’, also referred to as ‘the anti-censorship bill’ at the end of 2021.¹¹¹ The controversial draft law was set to prevent social media platforms from including content removal and account suspension measures within their content standards, limiting their ability to make their own content moderation decisions. However, as of November 2022, the law had still yet to be enacted. Additionally, the government of Poland was also set to amend the ‘National Cybersecurity System Act’. The original proposal would have given the government powers to block websites and enforce content removal via a specific security order. However, the act has failed to pass several times and, as of October 2022, the government has published the eighth version of the proposal.¹¹²



Tanzania

In March 2022, Tanzania’s Minister for [Information, Communication and Information Technology amended the Electronic and Postal Communications \(Online Content\) Regulations 2020](#). Following backlash to the 2020 regulations, the 2022 amendments were focused on narrowing the breadth of companies of service providers in scope, namely removing any reference to internet service providers and re-categorising the types of content in scope. It also repealed specific regulations, including the complete removal of regulation 13, which stipulated several actions on behalf of internet cafés, including the ability to restrict access to prohibited content, including terrorist content, through filtering.

¹⁰⁸ The Online Regulation Series (2021), [The Online Regulation Series 2.0 – Kazakhstan](#)

¹⁰⁹ Freedom House (2022), [Freedom on the Net 2022 – Kazakhstan](#)

¹¹⁰ Tech Against Terrorism (2021), [The Online Regulation Series 2.0 – Poland](#)

¹¹¹ Ibid

¹¹² Wachowska Agnieszka (2022), [A new version of the National Cybersecurity System Act close to being passed](#), TKP



Philippines

Despite having a high internet penetration rate, the Philippines has a limited online regulation framework. As highlighted in analysis within our first edition of the Online Regulation Series Handbook,¹¹³ the government of the Philippines have previously discussed the possibility of regulating social media through the extension of the Anti-Terrorism Act, but this was met with sharp criticism and was never actioned. However, in February 2022, the Congress of the Philippines' passed the Sim Card Registration Bill, which was designed to counter online abuse.¹¹⁴ The bill stipulated that SIM cards must be registered with telecommunications operators before purchase and that users of social media platforms must provide their real names to set up an account.¹¹⁵

Additionally, if users' were to use fake names upon registration, the draft law outlines financial and criminal penalties, namely a fine of 200,000 pesos and a minimum minimum six-year prison sentence.¹¹⁶ This draft law was met with significant criticism from civil society groups and in April 2022, the former president Duterte used his vetoing powers to reject the bill, citing opportunities for surveillance and infringements of users' human rights.



Ireland

In January 2022, the Irish government published the [Online Safety and Media Regulation \(OSMR\) Bill](#), and tabled it for the consideration of Seanad Éireann. The Bill, first introduced in December 2020, aims to tackle "harmful" content online and to align Ireland with the EU's Audiovisual Media Services Directive of 2018. As the Bill passed through the Irish upper house, a key debate among policymakers was whether the legislation should include an individual complaints mechanism, with an amendment being included for individual complaints around certain categories once the service provider's complaint mechanism has been exhausted.

The government will also amend the legislation to widen the scope of the Coimisiún na Meán (the new media regulator) to become the designated competent authority under the EU Terrorist Content Online Regulation, which will include overseeing the implementation of specific measures by service providers and the potential imposition of administrative fines.¹¹⁷ Other notable amendments introduced to the OSMR Bill include the appointment of an online safety commissioner¹¹⁸ within the Coimisiún na Meán, and the inclusion of 'cyber-flashing'¹¹⁹ as a further category of offence-specific harmful content online. The most recent [amendments](#) were tabled in late October 2022 with the Bill having completed Dáil Éireann, Third Stage, and was enacted at the end of 2022.¹²⁰

¹¹³ Tech Against Terrorism (2021), [The Online Regulation Series Handbook](#), p.70

¹¹⁴ Morales Neil Jerome (2022), [Philippines passes law to tackle anonymous social media abuse](#), Reuters

¹¹⁵ Maros Christia Marie (2022), [SIM registration bill just a step away from becoming law](#), Inquirer.net

¹¹⁶ Freedom House (2022), [Freedom on the Net 2022 – Philippines](#)

¹¹⁷ Law Society Gazette (2022), [Online watchdog to enforce terrorist-content rules](#)

¹¹⁸ Irish Legal News (2022), [Online safety commissioner to be appointed under legislation](#)

¹¹⁹ The Journal (2023), ['Cyber-flashing' to be tackled under new online safety legislation](#)

¹²⁰ Lexology (2022), [In brief: media law and regulation in Ireland](#)

Nigeria

At the time of publishing the Online Regulation Series 2.0 in November 2021, the Nigerian parliament was reviewing the controversial “Protection From Internet Falsehood and Manipulation Bill” or “social media bill”, first introduced in 2019 with the aim to curb the spread of “falsehood and fake news” in Nigeria. In June 2022, the National Information Technology Development Agency (NITDA), a federal government agency, issued a draft [Code of Practice for Interactive Computer Service Platforms/ Internet Intermediaries](#) (the “Code”). With the “social media bill” side-lined, the draft Code aims to set out best practices of Interactive Computer Service Platforms/Intermediaries as well as setting out measures “to combat online harms such as disinformation and misinformation.”

The provisions in the Code are extensive and reflect some concerning global regulatory trends highlighted by Tech Against Terrorism. These include mandating a local physical presence for large platforms, mandatory annual compliance reports, and a 24-hour removal deadline for unlawful or prohibited content. Additionally, the draft Code attempts to define harmful content which includes unlawful content (violating existing Nigerian laws) and prohibited content (“objectionable” on the grounds of public interest, morality, order, security, peace). To help enforce these obligations, platforms will have to provide dedicated complaint channels where authorised government agencies or users can lodge complaints against illegal or prohibited content, with penalties for non-compliance with the Code including fines and imprisonment.¹²¹ Upon a court order, tech platforms will have to submit information when requested by an authorised government agency to assist with combatting cybercrimes.

NITDA’s draft Code has faced widespread criticism from civil society organisations in Nigeria, and has been labelled a “reincarnation” of the unpopular “social media bill”.¹²² The most notable concern relates to the serious threat posed to freedom of expression (guaranteed by local¹²³ and international law¹²⁴) by the definition of harmful content which is vague and open to abuse by the government.¹²⁵ This risk is exacerbated by the mandated short removal deadline and dedicated channel for government complaints, incentivising overbroad censorship of online speech.

Additionally, despite being called a “Code of Practice”, Part 6 creates civil or criminal offences under the NITDA Act 2007 for noncompliance, with platforms and their staff liable for disciplinary measures and prosecution. Placing legal liability on platforms and their staff for user-generated content exposes platforms to additional risks that incentivise over removal of content and can lead to their withdrawal from the jurisdiction, further impacting the right to access information online. There are also concerns about the right to privacy being threatened by the requirement for social media platform to assist investigations by providing information and clause V.10 which requires hosts to “trace, expose, penalize, and close” accounts spreading mis- or disinformation.¹²⁶

¹²¹ Dentons Acas Law (2022), [NITDA Draft Code for Interactive Computer Service Platforms and Internet Intermediaries](#)

¹²² Eze Gabriel (2022), [What Digital Platforms and Intermediaries Should Know about NITDA’s Code of Practice for Nigeria’s Digital Space](#), Infusion Lawyers

¹²³ Section 39 of Nigerian Constitution

¹²⁴ Article 19 of Universal Declaration of Human Rights

¹²⁵ Amnesty International (2022), [Nigeria: NITDA Code of Practice must comply with International Human Rights Law](#)

¹²⁶ Ibid



Turkey

On 13 October 2022, Turkey's parliament adopted a "disinformation law" with the declared aim of tackling disinformation online. However, its opponents have dubbed it the "censorship law"¹²⁷, highlighting the law's enabling of undue governmental control over online speech. Article 29 has received particular attention in this regard, prescribing three years imprisonment for anyone who disseminates "false information" with the intent to "instigate fear, panic" or "endanger the country's security, public order and general health of society."¹²⁸ Tech companies have been made "directly responsible" for policing "unlawful" content, and face fines of up to 3% of their global revenue for non-compliance, as well as advertising bans.¹²⁹

Additional measures have been introduced to ensure strict compliance with the law's provisions, including a four-hour time limit for platforms to remove disinformation following a court order and the threat of 'throttling', which involves reducing a platform's bandwidth to effectively block it in the country. User privacy is also under threat as social media companies are required to hand over the personal details of users spreading "fake news"¹³⁰ as well as information relating to certain crimes including child sexual abuse material (CSAM), disinformation and state secrets, upon request by government.¹³¹ Information about algorithms, promoted or demoted content, advertisement and transparency policies should also be made available to the Information and Communication Technologies Authority (ICTA) upon request.

The new legislation, which was voted through Parliament by President Erdogan's ruling AK Party (AKP) and its allies MHP, has raised widespread concerns amongst opposition lawmakers, media freedom advocates, and legal scholars, especially in light of the upcoming presidential and parliamentary elections. According to Amnesty International, "the law's vaguely defined provisions facilitate further the prosecution of those who allegedly publicly disseminate 'false information' and could see people facing jail terms of up to three years merely for a retweet."¹³²

The law gives prosecutors extensive powers to determine what the truth is and according to the Council of Europe's legal advisory body, the Venice Commission, will likely lead to self-censorship.¹³³ Some regulatory and legal analysts have warned that large social media companies are unlikely to fully abide by the new law due to global data privacy standards and the dangerous precedent it could set.¹³⁴

¹²⁷ Adal Hikmet (2022), [Disinformation law 'final stage of AKP's media restrictions'](#)

¹²⁸ York Jillian (2022), [Turkey's New Disinformation Law Spells Trouble For Free Expression](#), Electronic Frontier Foundation

¹²⁹ Duvar.English (2022), [Social media companies might resist Turkey's new 'disinformation law'](#)

¹³⁰ Hubbard Ben and Timur Safak (2022), [Turkey Allows Jail Terms for What It Deems 'Fake News'](#), The New York Times

¹³¹ York Jillian (2022), [Turkey's New Disinformation Law Spells Trouble For Free Expression](#), Electronic Frontier Foundation

¹³² Amnesty International (2022), [Turkey: "Dark day for online free expression" as new 'disinformation law' is passed](#)

¹³³ Council of Europe (2022), [Urgent Joint Opinion of the Venice Commission and the Directorate General of Human Rights and Rule of Law \(DGI\) of the Council of Europe on the draft amendments to the Penal Code regarding the provision on "false or misleading information"](#)

¹³⁴ Duvar.English (2022), [Social media companies might resist Turkey's new 'disinformation law'](#)



Pakistan

Pakistan passed the Citizen Protection (Against Online Harm) Rules in 2020. The Social Media Rules became effective in 2021, and in the same year the Pakistani government announced that it would revise them. The Rules and its revisions, which included a requirement for platforms to comply with emergency requests from the Pakistan Telecommunication Authority (PTA) within 12 hours, were met with widespread criticisms by industry groups and digital rights advocates in Pakistan.

Tech Against Terrorism last provided a short update on Pakistan’s Social Media Rules in the Online Regulation Series 2.0 published in November 2021. At the time, the Islamabad High Court (IHC) was to review the Social Media Rules according to National Standards. In 2022, the IHC ruled that the Social Media Rules were to be referred back to the National Assembly for review and for the inclusion of amendments safeguarding freedom of expression. This decision was partly based on the change of government in Pakistan, as the political party contesting the rules is now in power in the country.¹³⁵

In September 2022, the Ministry of Information Technology and Telecommunication announced that it had established a committee to review the rules. The first meeting was announced to take place on 8 September 2022. The Ministry also stated that the revised Social Media rules would be announced at the end of October. However, at the end of November 2022 it was unclear when they would be released.



Uganda

On October 13 2022, Ugandan President Yoweri Museveni signed into law new legislation that criminalises certain types of internet activity, tightening restrictions brought by the 2011 Computer Misuse Act, which aims to prevent the abuse of information systems.¹³⁶ Clause 6 of the law criminalises the publishing, distributing or sharing of information prohibited under Uganda’s laws with punishment of imprisonment for up to five years, a fine of up to UGX 10 million (USD 2,619), or both.¹³⁷ Another provision stipulates prison sentences of up to ten years in some cases for offenses related to the transmission of information about a person without their consent.¹³⁸

Despite being justified as a necessary measure to tackle the sharing of “unsolicited, false, malicious, hateful and unwarranted information”¹³⁹ on social media platforms, press freedom groups are concerned the law provides the government with a tool to target critical voices and punish independent journalists.¹⁴⁰

¹³⁵ The Rules had been passed under the government of the Pakistan Tehreek-i-Insaf (PTI), with the petition for the IHC to review them filled by members of Pakistan Peoples Party, now in power.

¹³⁶ Valentiniy Anna (2022), [Uganda: New law tightens restrictions on internet use](#), International Press Institute

¹³⁷ Africa News (2022), [Uganda: New law imposes restrictions on use of the internet](#)

¹³⁸ Muhumuza Rodney (2022), [New Law in Uganda Imposes Restrictions on Use of Internet](#), CIPESA

¹³⁹ Muhumuza Rodney (2022), [New Law in Uganda Imposes Restrictions on Use of Internet](#), CIPESA

¹⁴⁰ Valentiniy Anna (2022), [Uganda: New law tightens restrictions on internet use](#), International Press Institute



Russia

Following Russia's invasion of Ukraine in February 2022, the government has operationalised its recent online regulatory framework to tighten its control over the internet, especially laws which expedite the blocking of online content prohibited under Russian law. The Sovereign Internet Law, comprising 2019 amendments to the [IT Law](#), has been increasingly utilised to block access to websites including independent Russian news websites, Ukrainian domains, large tech companies, and foreign news sites. Since the start of the war, more than 2,384 sites have been blocked within Russia, including Facebook, Instagram and Twitter.¹⁴¹

Meta has even been added to a list of terrorist and extremist organisations alongside groups such as the Taliban, making it a criminal offence for Russians to use Meta platforms Instagram and Facebook.¹⁴² The Russian regulator, Roskomnadzor, has also targeted VPNs, using the 'Extremist Websites Blocking Law' and 'VPN Law' to remove hundreds of thousands of VPN-related links from Google and utilise local Internet Service Providers (ISPs) to interfere with VPN connections.¹⁴³

The [Foreign Social Media Law](#), enacted in January 2021 to create penalties for foreign social media companies for failing to restrict access to content deemed illegal by Russian law, has been deployed to pursue fines against large foreign tech companies. In July 2022, Google was ordered to pay \$370 million by a Russian court for its "repeated failure" to remove "prohibited content" deemed "fake"¹⁴⁴, resulting in its Russian subsidiary filing for bankruptcy following seizure of its bank accounts.¹⁴⁵ This included content discrediting Russian armed forces, promoting "extremism", or inciting people to join public protests. Meanwhile, Apple was fined 2 million rubles (\$34,000) for violating Russia's data storage law, by refusing to store the personal data of Russians users on servers in Russia.¹⁴⁶

Freedom of expression concerns have arisen around reports that Roskomnadzor is also acting as a surveillance body to monitor and report anti-government users online.¹⁴⁷ According to leaked documents, Roskomnadzor monitors websites, social media, and news outlets for 'anti-government' accounts who are then investigated, with users' details passed on to security services.¹⁴⁸

According to the former UN Special Rapporteur David Kaye on the promotion and protection of the right to freedom of opinion and expression, David Kaye, Russian state actions since the invasion "have demonstrated the coercive power of the state over online expression, privacy, and public protest", highlighting a dependence on "censorship by its media regulator, and legal and extralegal demands against internet platforms."¹⁴⁹ As noted in our [previous analysis](#) of Russia's online regulatory framework,¹⁵⁰ the practices of website blocking and throttling, which the Russian authorities have increased their technical capability and deployment of over the past year,¹⁵¹ contravene international human rights law.

¹⁴¹ Migliano Simon and Woodhams Samuel (2023), [Websites Blocked in Russia Since Ukraine Invasion](#)

¹⁴² Tidy Joe (2022), Russia confirms Meta's designation as extremist, BBC

¹⁴³ Maxwell Andy (2022), [New VPN Crackdown Underway in Russia. Government Confirms](#), Torrent Freak

¹⁴⁴ Belanger Ashley (2022), [Russia fines Google \\$370M for refusing to bend to Putin's war propaganda](#), Ars Technica

¹⁴⁵ Ciacca Chris (2022), Google's [Russian subsidiary files lawsuits against Russian bailiffs: report](#), Seeking Alpha

¹⁴⁶ The Independent (2022), [Russia fines Apple for violating data storage law](#)

¹⁴⁷ Mozur Paul, Satariano Adam, Krolik Aaron, Aufrichtig Aliza (2022), [They Are Watching: Inside Russia's Vast Surveillance State](#), The New York Times

¹⁴⁸ Sherman Justin (2022), [Russia's Internet Censor is Also a Surveillance Machine](#), Council on Foreign Relations

¹⁴⁹ Kaye David (2022), [Online Propaganda, Censorship and Human Rights in Russia's War Against Reality](#), American Journal of International Law

¹⁵⁰ Tech Against Terrorism (2021), [The Online Regulation Series 2.0 – Russia](#).

¹⁵¹ Sherman Justin (2022), [Russia's Internet Censor is Also a Surveillance Machine](#), Council on Foreign Relations





SECTION 3.

CROSS-SECTOR POLICY INITIATIVES

In this section, we welcome our colleagues from the Christchurch Call Advisory Network, to provide an excerpt from the pilot project on “Evaluating the Impact of Government and Company Commitments Under the Christchurch Call to Action”.



EVALUATING THE IMPACT OF GOVERNMENT AND COMPANY COMMITMENTS UNDER THE CHRISTCHURCH CALL TO ACTION

A Pilot Project of the Christchurch Call Advisory Network

To mark the third year since the Christchurch Call to Action, the Christchurch Call Advisory Network (CCAN) embarked on a first-ever independent evaluation of the work done by supporting governments and companies (“supporters”) to further the Call. Through this pilot evaluation, the CCAN engaged with key signatories (a selective sample of the CCAN supporters) to understand how the commitments of these governments and companies under the Call had shaped their approaches to curbing the spread of terrorist and violent extremist content in a manner consistent with human rights and a free, open, and secure internet. In addition, CCAN’s pilot evaluation examined ways in which supporting governments and companies engaged in multi-stakeholder discussion and policy development within the broader Call Community. This section will summarize the Call’s commitments, its impact on supporters of the Call and where supporters can do more to uphold and implement key tenets of the Call’s commitments.

The Christchurch Call to Action was established in May 2019, two months after a terrorist attack in a mosque in Christchurch, New Zealand killed 51 people and injured 50 more. Aotearoa New Zealand Prime Minister Jacinda Ardern and French President Emmanuel Macron brought together Heads of State and leaders from the technology sector to adopt the Christchurch Call. Seventy governments and companies are now signatories of the Call.¹⁵² The Call also includes the Christchurch Call Advisory Network, whose more than 68 members from civil society include advocates, activists, academics, and technical experts who volunteer to provide advice to governments and online service providers and engage in the broader work of the Call.

Key commitments of the Call include countering the drivers of terrorism and violent extremism through legislative, programmatic, and collaborative efforts; effectively enforcing applicable laws; supporting frameworks across media, industry, and other sectors to ensure a coordinated response to countering the spread of terrorist and violent extremist content; reviewing the operation of algorithms and other processes which may drive users towards this dangerous content; and engaging with civil society and multi-stakeholder arenas.¹⁵³

¹⁵² See the full list of supporters on the Christchurch Call website: <https://www.christchurchcall.com/our-community/countries-and-states/>.

¹⁵³ The full Christchurch Call text on the Call’s website: <https://www.christchurchcall.com/about/christchurch-call-text/>

Methodology


In CCAN's evaluation of supporting governments and companies, we first selected a subset of the Call's commitments to focus our assessment on some overarching themes of the Call: transparency, human rights due diligence, civil society engagement, and cross-Call collaboration. CCAN also selected a small sample of the Call's supporting governments and companies to include in this first evaluation. We chose six governments — Aotearoa New Zealand, France, Australia, Canada, United Kingdom, and India—and four companies—Microsoft, Meta, Twitter and Google. We chose these signatories based on their role as leaders of the Call (in the case of Aotearoa New Zealand and France), the longevity of their support for the Call, and our internal capacity to conduct this analysis (such as familiarity with language, legal systems, and access to resources). Our evaluation was conducted through two workstreams: (1) direct engagement with the governments and companies that we evaluated and (2) independent research of publicly available information.

Findings

Overall, the supporting governments and companies we evaluated **were explicit about their shared commitment to the Call's principles. Each of them clearly communicated through their public channels and media that they had joined the Call and were focused on curbing the spread of terrorist and violent extremist content online.** Some signatories had taken legislative action in relation to terrorist material online and the broadcasting of terrorist livestreams. Some also established independent authorities to provide oversight of these issues. This supports the conclusion that the Call has played a significant role in highlighting the urgency of this problem and the importance of coordination in tackling it.

In contrast, it was much **harder to find evidence that supporters had implemented their commitments under the Call beyond declarations of intent to do so.** If work was undertaken in response to the Call, it was rarely identified as such, making measurement of the Call's impact difficult. Moreover, without direct engagement with representatives of supporting governments and companies, it was difficult to ascertain which agency or department (often multiple) were responsible for carrying out the commitments under the Call and which served as the point of contact. Similarly, information about initiatives launched in concert with the Call, any human rights oversight and assessments conducted of existing applicable laws and policies, and instances of engagement with civil society, were scarcely documented.

Supporting governments and companies also note that an absence of agreed-upon shared definitions of terms like 'terrorist content' made implementing the Call's commitments difficult. Additionally, limited research conducted on the quantity and quality of violence-inducing content on member services like LinkedIn and Bing, and Voice over Internet Protocol (VOIP) services like Skype and Zoom serve as barriers to completing a comprehensive and conclusive evaluation. More research is needed to understand the scope of terrorist and violent extremist content on all platforms, and more transparency from platforms is needed to understand the effectiveness of existing policies and protocols in taking action on violative content. The Call can play a role in convening discussions around shared definitions and what it means to "effectively" curb the spread of terrorist and violent extremist content as well as promote further research into this space.



Finally, there was an absence of oversight mechanisms to hold Call supporters accountable to their commitments and guide their approaches. Supporters pursued a wide variety of approaches to address the proliferation of terrorist and violent extremist content online; some of these approaches were difficult to implement and others had unintended negative consequences, including further stigmatizing affected communities and hampering individuals' ability to access information online. Both of these outcomes run counter to the Call's commitments to "ensure effective enforcement of applicable laws" and consider "regulatory or policy measures consistent with a free, open and secure internet and international human rights law". To combat these issues, the Call could play a stronger role here, as the global coordination tool that it is. The Call structure can create opportunities for supporters to present forthcoming initiatives intended to adopt the Call's principles, and the Call community, including CCAN members, offer a range of human rights, legal, and anti-violence expertise and consider trade offs present in specific regulatory, policy, transparency, and engagement initiatives.



tech against terrorism



BIBLIOGRAPHY



BIBLIOGRAPHY

Access Now (2017), [Access Now submission to the Universal Periodic Review: UAE, Third Cycle](#).

Access Now (2022), [Civil society calls on Indian government to withdraw amendments to IT Rules](#)

Access Now (2022), [India's Draft Telecommunication Bill must be revamped to protect human rights](#)

Access Now (2022), [Global coalition of NGOs urge Indonesia to repeal censorship regulations](#)

Adal Hikmet (2022), [Disinformation law 'final stage of AKP's media restrictions'](#)

Africa News (2022), [Uganda: New law imposes restrictions on use of the internet](#)

Article19 (2020), [Kyrgyzstan: Draft Law on Countering Terrorism](#)

Article19 (2021), [Kyrgyzstan: Report on freedom of expression and 'extremism'](#)

Article19 (2022), [United Arab Emirates: New cybercrime and anti-rumour law violates rights](#).

Allen Asha (2022), [The Digital Services Act: Political Agreement Reached, Long Road Ahead Awaits](#), Center for Democracy & Technology

Amnesty International (2021), [UAE: Nearly a decade of unjust imprisonment for 'UAE-94' dissidents](#)

Amnesty International (2022), [Nigeria: NITDA Code of Practice must comply with International Human Rights Law](#)

Amnesty International (2022), [Turkey: "Dark day for online free expression" as new 'disinformation law' is passed](#)

Anderson Alex (2022), [Addressing Health Misinformation](#), Discord

BBC (2020), [Vienna shooting: What we know about 'Islamist terror' attack](#)

Belanger Ashley (2022), [Russia fines Google \\$370M for refusing to bend to Putin's war propaganda](#), Ars Technica

Bonifacic Igor (2020), [Instagram starts labeling 'state-controlled media' accounts and posts](#), Engadget

Bonifacic Igor (2022), [Indonesia blocks Steam, PayPal and other services over missed regulatory deadline](#), Engadget.

Caster Michael (2022), Internet Freedom in Indonesia is Teetering on a Razor's Edge, The Diplomat

Christchurch Call to Action (2022), [Christchurch Call Initiative on Algorithmic Outcomes](#)

Ciacca Chris (2022), [Google’s Russian subsidiary files lawsuits against Russian bailiffs: report](#), Seeking Alpha

Council of Europe (2022), [Urgent Joint Opinion of the Venice Commission and the Directorate General of Human Rights and Rule of Law \(DGI\) of the Council of Europe on the draft amendments to the Penal Code regarding the provision on “false or misleading information”](#)

Counter Extremism Project (2022), [Austria: Extremism and Terrorism](#).

Curtis Barnes, Tom Barraclough, and Robins Allyn(2022) [Platforms Are Testing Self-Regulation in New Zealand. It Needs a Lot of Work.](#), Lawfare

David Kaye and Jason Pielemeier (2020), [The Right Way to Regulate Digital Harms](#), Project Syndicate

Dentons Acas Law (2022), [NITDA Draft Code for Interactive Computer Service Platforms and Internet Intermediaries](#)

Duvar.English (2022), [Social media companies might resist Turkey’s new ‘disinformation law’](#)

Electronic Frontier Foundation, [Manilla Principles on Intermediary Liability](#)

Eze Gabriel (2022), [What Digital Platforms and Intermediaries Should Know about NITDA’s Code of Practice for Nigeria’s Digital Space](#), Infusion Lawyers

Kashyap Hemant (2022), [Digital India Act Will Monitor Social Media, Metaverse, OTT Platforms: Report](#), Inc42

Freedom House (2021), [Freedom on the Net 2021 – Kyrgyzstan](#)

Freedom House (2022), [Freedom on the Net 2022 – Kyrgyzstan](#)

Freedom House (2022), [Freedom on the Net 2022 – Kazakhstan](#)

Freedom House (2022), [Freedom on the Net 2022 – Brazil](#)

Freedom House (2022), [Freedom on the Net 2022 – Philippines](#)

Freedom House (2022), [Freedom in the World 2022: The United Arab Emirates](#)

Goujard Clothilde (2022), [Twitter to take down RT, Sputnik after EU sanctions](#), Politico.eu

Hubbard Ben and Timur Safak (2022), [Turkey Allows Jail Terms for What It Deems ‘Fake News’](#), The New York Times

Human Rights Watch (2018), [‘We Live in Constant Fear’ - Possession of Extremist Material in Kyrgyzstan](#)



Human Rights Watch (2022), [UAE: Sweeping Legal ‘Reforms’ Deepen Repression.](#)

International Centre for Justice and Human Rights (2016), [ICJHR submission to the Office of the UN High Commissioner for Human Rights on the negative effects of terrorism on the enjoyment of all human rights and fundamental freedoms in the United Arab Emirates](#)

Internet Society (2022), [Internet Impact Brief: Proposals to Regulate Content Moderation on Social Media Platforms in Brazil](#)

Irish Legal News (2022), [Online safety commissioner to be appointed under legislation](#)

Kaye David (2019) [Promotion and protection of the right to freedom of opinion and expression : note by the Secretary-General](#), UN. Human Rights Council. Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression

Kaye David (2022), [Online Propaganda, Censorship and Human Rights in Russia’s War Against Reality](#), American Journal of International Law

Khaleej Times (2022), [UAE: 17 categories of online content, websites that are blocked in the emirates](#)

Law Society Gazette (2022), [Online watchdog to enforce terrorist-content rules](#)

Lexology (2022), [In brief: media law and regulation in Ireland](#)

Maros Christia Marie (2022), [SIM registration bill just a step away from becoming law](#), Inquirer.net

Maxwell Andy (2022), [New VPN Crackdown Underway in Russia, Government Confirms](#), Torrent Freak

Menn Joseph (2021), [WhatsApp sues India govt. says new media rules mean end to privacy](#), Reuters (republished in the Financial Post)

Migliano Simon and Woodhams Samuel (2023), [Websites Blocked in Russia Since Ukraine Invasion](#)

Morales Neil Jerome (2022), [Philippines passes law to tackle anonymous social media abuse](#), Reuters

Mozur Paul, Satariano Adam, Krolik Aaron, Aufrichtig Aliza (2022), [They Are Watching’: Inside Russia’s Vast Surveillance State](#), The New York Times

Muhumuza Rodney (2022), [New Law in Uganda Imposes Restrictions on Use of Internet](#), CIPESA

Newton Casey (2022), [Russia’s propaganda network gets deplatformed](#), Platformer

Noueihed Lin (2008), [UAE’s du starts restricting Internet access](#), Reuters.

O’Sullivan Donie (2022), [Twitter is no longer enforcing its Covid misinformation policy](#), CNN

Porutiu Theodor (2022), [Censorship in the UAE: How to Get Around it](#), VPN Overview.

Potkin Fanny and Sulaiman Stefano (2022), [Indonesia preparing tough new curbs for online platforms](#), Reuters.

[The Santa Clara Principles On Transparency and Accountability in Content Moderation](#)

Sawers Paul (2022) [Big Tech's push to self-regulate harmful content in New Zealand is 'weak attempt to preempt regulation', critics say](#), TechCrunch

Scott Mark (2022), [Telegram bans Russian state media after pressure from Europe](#), Politico.eu

Sherman Justin (2022), [Russia's Internet Censor is Also a Surveillance Machine](#), Council on Foreign Relations

Singh Manish (2022), [American internet giants seek changes to India's proposed new IT rules](#), TechCrunch

Windwehr Svea and York Jillian (2020), [Thank You For Your Transparency Report, Here's Everything That's Missing](#), The Electronic Frontier Foundation.

Tham Irene (2022), [New rules to make social media firms accountable for online harms](#), The Straits Times

The Independent (2022), [Russia fines Apple for violating data storage law](#)

The Journal (2023), ['Cyber-flashing' to be tackled under new online safety legislation](#)

Thompson Peter and Michael Daubs (2021) [Executive Digest: International Regulatory Frameworks for Online Content](#), p. 9

Tidy Joe (2022), [Russia confirms Meta's designation as extremist](#), BBC

Twitch, [Preventing Harmful Misinformation Actors on Twitch](#)

United Nations General Assembly [Resolution 68/167](#)

United Nations Office of the High Commissioner for Human Rights (2017), [Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression](#)

Valentiniy Anna (2022), [Uganda: New law tightens restrictions on internet use](#), International Press Institute

Wachowska Agnieszka (2022), [A new version of the National Cybersecurity System Act close to being passed](#), TKP

Yeo Amanda (2022), [Reddit has quarantined r/Russia due to misinformation](#), Mashable

York Jillian (2022), [Turkey's New Disinformation Law Spells Trouble For Free Expression](#), Electronic Frontier Foundation



Tech Against Terrorism resources

Tech Against Terrorism (2021), [The Online Regulation Series Handbook](#).

Tech Against Terrorism (2021), [Transparency Reporting Guidelines on Online Counterterrorism Efforts](#).

Tech Against Terrorism (2022), [The Online Regulation Series 2.0 – Summary](#)

Tech Against Terrorism (2023), [The Tech Against Terrorism Podcast – “Sanitising Extremism: ‘Borderline Content’ and Antisemitism Online”](#).

Government resources

Government of Austria (2012), [Section 278f of the Criminal Code Instructions for committing a terrorist offense](#).

Government of Austria (2012), [Section 278b of the Criminal Code Terrorist Association](#).

Government of Austria (2012), [Section 282a of the Criminal Code, Incitement to commit terrorist offenses and approval of terrorist offenses](#).

Bundesministerium Inneres, [Reporting Office for Extremism and Terrorism](#).

RTR, [The Organisation](#).

Government of India, <https://egazette.nic.in/WriteReadData/2022/239919.pdf>

Government of India, https://www.mha.gov.in/sites/default/files/EighthSchedule_19052017.pdf

Government of India, <https://dot.gov.in/sites/default/files/Explanatory%20Note%20to%20the%20draft%20Indian%20Telecommunication%20Bill%2C%202022.pdf>

EUR-Lex, [Primacy of EU law](#)

European Commission (2022), [Digital Services Act: EU’s landmark rules for online platforms enter into force](#)


Council of the European Union (2022), [EU imposes sanctions on state-owned outlets RT/ Russia Today and Sputnik’s broadcasting in the EU](#)

European Internet Forum (2023), [The Revised EU Crisis Protocol: Responding To Terrorist Content Online](#).

New Zealand Cabinet (2021), [Proactive release of Cabinet material about the initiation of the media content regulatory review](#)

New Zealand Department of Internal Affairs, [Objectionable and Restricted Material](#)

Embassy of the United Arab Emirates Washington DC, [Counterterrorism](#).



Government of Brazil – Ministry of Tourism, <https://static.poder360.com.br/2021/06/minuta-decreto-mudanca-marco-civil-internet.pdf>

Government of the United Kingdom – Department for Digital Culture, Media & Sport (2022), [New protections for children and free speech added to internet laws](#)

Indonesia, Ministry of Communication and Informatics (2021), [Regulation of the Minister of Communication and Informatics Number 10 of 2021 concerning Amendments to the Regulation of the Minister of Communication and Informatics Number 5 of 2020 Concerning Private Electronic System Operators](#)



This work is licensed under a Creative Commons Attribution – Non-Commercial – NoDerivatives 4.0 International Licence. You are free to reference and cite this publication so long as you cite the source of this report: Tech Against Terrorism. For more information, see <http://creativecommons.org/licenses/by-nc-nd/4.0/>